(4)

Systems
Optimization
Laboratory

FORMULATION AND SOLUTION OF ECONOMIC EQUILIBRIUM PROBLEMS

John C. STONE

TECHNICAL REPORT SOL 88-7

April 1988

Department of Operations Research
Stanford University
Stanford, CA 94305

(4)

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022

FORMULATION AND SOLUTION OF ECONOMIC EQUILIBRIUM PROBLEMS

John C. STONE

TECHNICAL REPORT SOL 88-7
April 1988

Research and reproduction of this report were partially supported by National Science Foundation Grants DMS-8420623, SES-8518662, and ECS-8617905; U.S. Department of Energy Grant DE-FG03-87-ER25028; Office of Naval Research Contract N00014-85-K-0343; and the Center for Economic Policy Research at Stanford University.

ACKNOWLEDGEMENTS

It is more than just proper form to say that an effort of this magnitude could not have been completed without the help and support of friends, colleagues, and family. It is absolutely true. I would like for a number of people to know how much I appreciate their support.

Thank you, George Dantzig, for supporting me these many years, both in terms of employment and in giving me enough rope to do things my own way — even when I wrapped that rope around my own neck.

Thank you, George, Richard Cottle, and Curtis Eaves, for your commentary on the first draft and for helping me over some of the rough technical spots along the way.

Thank you, Alan Manne, for helping to get me (and others in this Department) involved with general equilibrium modeling in the first place. Thank you, Alan and Jackie Manne, for caring enough to start nagging me to get finished.

Thank you, Patrick McAllister, for taking the time to read through the first draft so carefully and for the many discussions we have had about the subtleties of equilibrium problems.

Thank you, Michael Saunders; your patience knows no bounds. I honestly do not know where I would be without your many hours of help concerning both my using (and abusing) MINOS for the computations and my specializing LaTeX for preparing this document.

Thank you, Thomas Rutherford, Olvi Mangasarian, Philip Gill, and Margaret Wright, for numerous fruitful discussions about this research.

Thank you, Gail Stein and Audrey Stevenin, for your friendly support and your help with a million little details along the way.

Thank you, Charles and Mary Bess, my parents, for always making sure that it was possible for me to spend my years at Harvard and here at Stanford. Thank you also, I think, for doing whatever it was you did to make me the perfectionist I am today. Maybe now I can give it a rest.

Finally, thank you, Lupita, John, Elizabeth, and Christine, my Pinter family, for putting up with this cranky procrastinator these too many years that it has taken to finish this effort.

John Charles Stone
Stanford, California
March 1988

General equilibrium starts at home.
Lupita Arce Pinter

ABSTRACT

We develop and assess a number of equivalent mathematical formulations of the general equilibrium problem in economics. We begin with the traditional representation as a nonlinear complementarity problem and develop alternative representations as nonlinear optimization problems. All of our formulations depart from previous approaches by including an explicit linear equality for price normalization and a matched artificial variable which must be zero at any equilibrium solution. This structure has the theoretical and computational advantage that any linearization of the equilibrium problem has a feasible complementary solution. Moreover, under a reasonable rank condition, a basic complementary solution exists which can be successfully computed by Lemke's almost-complementary pivoting method.

We describe five general-purpose methods which can be applied to solving the equilibrium problem. The common feature of these methods is solving a sequence of linearized problems. We establish a number of equivalences between the methods, when applied to the equilibrium problem, and assess their local and global convergence properties in that context. An important new tool in this analysis is another problem formulation based on a differentiable exact penalty function. This formulation provides, perhaps for the first time, a rigorous means of evaluating the progress of a sequential method for computing an equilibrium solution. Our analysis reveals a basic theoretical dilemma in solving general equilibrium problems by these sequential methods. One group of methods produces iterates that converge from any starting point, but the sequence may converge to a non-equilibrium point. Another group produces iterates that may fail to converge, but successfully converging sequences do attain an equilibrium.

We perform a number of computational experiments on two small problems from the literature. The results show considerable variation in the solution times for the various methods, but all methods succeed in locating an equilibrium, even from poor starting points. This successful performance (in addition to that reported by other researchers) suggests that the kinds of general equilibrium models formulated in practice possess certain favorable computational properties that theoretical analysis has yet to discover.

CONTENTS

LIST OF FIGURES

1

INTRODUCTION

An important and powerful paradigm in economic theory is the notion of general equilibrium, that is, a set of prices (and corresponding quantities) which balance the supply and demand for all commodities. The term general equilibrium is used equally to refer to the market-clearing prices and quantities, themselves, as well as to the "invisible hand" mechanism by which markets are cleared. The concept of equilibrium is an old one, but it is also one for which current interest is both diverse and intense. The advent of the computer has given rise to a class of so-called *computable* general equilibrium models, which may be distinguished by the explicit intent to mathematically represent the equilibrium system, compute a solution of the system, and derive policy or other conclusions by comparing solutions across different model specifications. As well exemplified by the collection of papers in a recent *Mathematical Programming Study* [Mann 85], ever more sophisticated applications of computable general equilibrium models both stimulate and take advantage of improvements in computational procedures.

Against this backdrop, the present research is concerned with various aspects of the formulation and solution of general equilibrium problems. The key idea will be to define, compare, and exploit the properties of a number of different but equivalent mathematical formulations of the problem. While most of the results presented are theoretical in nature, our underlying interests and motivations are very much computational. These interests, coupled with some basic pragmatism, serve to establish some limits on the kinds of problems to be considered.

1.1 Scope of this research

First and foremost, our attention is limited to *general* equilibrium problems which require simultaneous determination of all prices (and quantities) in the economy. In contrast, *partial* equilibrium problems accept certain prices as given exogenously. This important difference has implications for the mathematical properties of the problems which, on the whole, make partial equilibrium models considerably easier to solve than general equilibrium models of comparable size and functional complexity. Partial equilibrium problems typically satisfy some contraction or monotonicity property which guarantees the convergence of an appropriate iterative method, examples of which are legion. Many successful approaches are specializations of the general iterative procedures for variational inequalities surveyed by Pang and Chan [PC 82]. An historically important example is the so-called PIES method [AH 82]. Also in this group are a variety of methods originally applied to network/traffic equilibria and more recently extended, in the work of Dafermos and Nagurney, to spatial equilibria and oligopolistic markets; see, for instance, [DN 84] and [Nag 87]. Viable approaches for specific large-scale models of energy markets have also been based on Newton-like methods [Phi 85] and on Gauss-Seidel iterations [Mur 83].

In sharp contrast, general equilibrium problems can almost never be shown to satisfy known sufficient conditions for the convergence of established iterative methods. (An important exception is noted below.) Consequently, procedures which perform efficiently on partial equilibrium models need not and typically do not perform especially well in a general equilibrium context. Conversely, while most of the findings and methods addressed in this research are also applicable to the partial equilibrium case, we would not expect the methods to be competitive with procedures that take explicit advantage of partial equilibrium properties.

In a general equilibrium context, there are three distinguishing features of a problem which determine not only how the real-world economy is stylized but also how difficult the problem proves to be in terms of computation:

1. departures from the pure competition paradigm
2. the representation of final consumer demand
3. the representation of production.

Simplifying assumptions in each of these areas define the class of general equilibrium problems to be considered in this research.

Departures from a purely competitive equilibrium market structure can take many forms. Taxes and subsidies on consumption or production activities effectively cause consumers and producers to see different prices for the same commodity. Price rigidities such as a minimum wage or regulated energy prices restrict the domain over which equilibrium prices can be sought, possibly preventing market-clearing altogether. Balance of payments constraints can also be included (somewhat loosely) under this heading. The work of Shoven and Whalley (e.g., [SW 73]) has spawned an entire generation of applied general equilibrium models specifically addressing taxes and other market interventions and imperfections. While such departures from pure competition are certainly characteristic of real-world economies, their representation in a model can present serious problems of both a theoretical and computational nature. Since the kind of computational analysis presented here has proved to be difficult enough even in the pure competition case, similar consideration of non-competitive structures must be left a subject for future research.

With respect to the representation of final consumer demand, we restrict our attention to twice continuously differentiable aggregate excess demand functions. As a theoretical matter, this assumes away situations in which a utility-maximizing choice occurs at a boundary of the space of consumption quantities (most typically, at a point where the consumption of some commodity is zero). The work of Mas-Colell [MasC 85] demonstrates that such situations are "rare" in a rigorous sense that need not be defined here. As a practical matter, virtually all demand functions which are actually used in applied equilibrium modeling satisfy this condition. What is ruled out are demand correspondences and piecewise-specified functions such as might be derived from corresponding linear or piecewise-specified utility functions for individual consumers. Two of the few examples of applications with nondifferentiable demand may be found in [DEG 79] and [MCW 80], wherein a number of consumer classes are each modeled explicitly by linear programs.

We also exclude from consideration the class of *integrable* demand functions, though in this case the reason is that models with such functions are relatively *easy* to solve. An aggregate demand function is said to be integrable if it can be derived from the optimality conditions for maximizing a corresponding utility function. In this case, it is well known that a general equilibrium solution can be computed (readily) by maximizing the utility function over the set of consumption levels which can be feasibly produced. As a theoretical matter, the conditions (on individual demands) required for the existence of an integrable aggregate demand function are quite restrictive. As a practical matter, integrable demand systems are often employed anyway, precisely because of the associated computational advantages. An example from our own work is the PILOT energy-economic model [SMD 87], in which the dynamic representation of production activity is so large that computation even with integrable demand functions is an expensive process. Indeed, it was precisely the desire to understand the computational consequences of abandoning the integrability assumption that initiated this research.

With respect to the representation of production, our focus is on models with activity analysis (or linear) production. This is a restrictive assumption, to be sure, although in principle any convex production structure can be approximated by a sufficiently large activity-analysis representation. Linear production figures prominently in early works on equilibrium and, in particular, in the pioneering computable formulations of Scarf [Sca 73]. It is also a principal feature of PILOT.

Happily, a model structure with (twice) continuously differentiable demand and linear production will also permit the incorporation of nonlinear production structures represented by (three times) continuously differentiable production (or profit) functions that reflect strictly decreasing returns to scale. This is accomplished by defining a composite demand function as the difference between consumer demand and the (nonlinear) net output from production, which in this case is uniquely determined as a function of prices. (The consumer demand component must be carefully defined so as to include the distribution of the positive net profits of production.) The composite demand function so defined possesses the same mathematical properties as the consumer demand functions employed throughout this research. The only restriction is the differentiability assumption, which in fact is satisfied by most of the nonlinear production or profit functions that are actually used in applied equilibrium modeling.

The production structures that are not immediately covered by our computational analysis are either nonconvex or involve nonlinear technologies with constant returns to scale. The latter omission is particularly unfortunate, given the theoretical importance and widespread use of such structures in equilibrium models. In the final chapter, we outline some very recent observations which suggest that most of the results obtained in this research can indeed be extended to the case of nonlinear production with constant returns.

After consideration of the above limitations of focus, we are left with an important class of competitive general equilibrium problems with continuously differentiable, nonintegrable demand and linear (or decreasing returns) production. Our theoretical analysis of solution algorithms for such problems will not explicitly address the issue of model size; nonetheless, our principal interest is with procedures which could be applied to medium- and large-scale

problems containing hundreds or thousands of commodities and activities. For this reason we will not address computational procedures based on homotopy or fixed-point methods. Such methods, as originated by Scarf [Sca 73] and more recently implemented by Broadie [Bro 85], are both general and robust, but their application to large-scale models is not currently feasible.

We also will not specifically analyze three related families of methods designed for the case in which nonintegrability of the final demand functions arises directly from aggregating known integrable demand functions for a manageable number of consumer classes. Indeed, the basic specification of consumer behavior is through the underlying direct utility functions. The methods of Dantzig, Eaves, and Gale [DEG 79] and of Ginsburgh and Waelbroeck [GW 81] seek a competitive general equilibrium by solving a sequence of optimization problems in which the objective function is a weighted sum of individual utilities. The method of Manne, Chao, and Wilson [MCW 80] solves utility-maximization problems for individual consumers so as to generate consumption bundles for use in a column generation procedure that seeks a combination of bundles consistent with aggregate supply-demand balances and the prices dual to those balances. While it may prove possible to analyze the convergence of these methods using the machinery of Chapter 5, such an analysis has not been performed as part of this research.

1.2 Tools of the trade

In order to fully characterize the relevant mathematical characteristics of a general equilibrium problem, Chapter 2 will provide a brief but reasonably self-contained discussion of the problem components and equilibrium conditions. Much of this material is likely to be familiar to the reader, but the repetition seems necessary in order to define the relevant context for the computational analysis that follows.

In contrast, it would be unrealistic to attempt to provide here a review of the many basic aspects of mathematical programming and complementarity theory that will be used extensively throughout the later chapters. We must in fact presume that the reader is more than familiar with these matters. Concepts to be employed include the standard formulations of mathematical programs (linear, quadratic, and general nonlinear) and of linear and nonlinear complementarity problems. We also employ the first- and second-order optimality conditions (both necessary and sufficient) and the associated fundamental concept of the Lagrangian function and Lagrange multipliers. Notions of basic solutions, uniqueness, and regularity prove to be important in several respects. In light of the breadth of the theories employed, we are also in no position to provide well-balanced citations for the many path-breaking contributions to this body of theory by Dantzig, Kuhn and Tucker, Cottle, and Lemke. A rigorous treatment of the fundamental concepts of mathematical programming, with appropriate references to the original contributions, can be found in the text of Avriel [Avr 76], for instance. A brief review of the basic concepts and origins of complementarity theory can be found in [Cot 76].

We also will be concerned with a variety of algorithms for solving mathematical programs and complementarity problems. The features of some of these algorithms will be reviewed as

they have direct bearing on our results. Reference will be made to Newton and quasi-Newton methods without elaboration. Penalty function concepts will be used freely. Aspects of convergence theory will be relevant, but there will be no detailed convergence proofs. A comprehensive survey of the state of the art in optimization algorithms (as of about 1980) is available in the text of Gill, Murray and Wright [GMW 81]. A brief review of more recent developments will soon appear in a forthcoming handbook of operations research [GMSW 87].

1.3 Outline of presentation

The balance of this document is organized as follows. First, we define some acronyms and notation that will be used recurrently in the later chapters. In Chapter 2 we build up the computable general equilibrium problem from its component parts, which entails describing the relevant mathematical properties of the demand and production structures, discussing the convention for normalizing prices, and delineating the mathematical conditions that define an equilibrium solution. In Chapter 3 we develop two alternative computational formulations of the general equilibrium problem. The first follows closely the statement of the problem in Chapter 2 and comprises a specially structured nonlinear complementarity problem. The second is a new development comprising a nonlinear program constructed to be equivalent to the complementarity problem. Both formulations depart from more traditional formulations by including both an explicit equality for price normalization and a matching artificial variable in the supply/demand balances.

In Chapter 4 we define five classes of methods for solving these equivalent formulations. The common feature of these methods is solving a sequence of subproblems defined by linearizing the aggregate demand functions. The methods may be distinguished by the type of linearization employed and the kind of solution sought relative to the linearized constraints. All of these methods are applicable to more general problems, but we demonstrate some interesting equivalences which arise when the methods are specifically applied to the equilibrium problem.

The local and global convergence properties of these methods are the subject of Chapter 5. We briefly survey known results and then examine the applicability of those results in the context of a general equilibrium problem. We develop an important new tool in understanding global convergence properties in the form of another equivalent formulation of the problem, which replaces the nonlinear constraints of the nonlinear programming formulation with a differentiable exact penalty function. This formulation provides, perhaps for the first time, a rigorous means of evaluating the progress of a sequential method for computing an equilibrium.

In Chapter 6 we demonstrate the value and power of our formulation of the problem with both the price normalization and the matching artificial variable. In particular, we prove that the normalization can readily be specified in such a way as to guarantee that any linearization of the nonlinear equilibrium problem has a feasible complementary solution. Moreover, under a reasonable rank condition (discussed in Chapter 4), a basic complementary solution exists which can be successfully computed by Lemke's almost-complementary

pivoting method.

In Chapter 7 we report the results of a number of computational experiments on two small problems from the literature. One has a unique equilibrium, while the other has three distinct equilibria. The results show considerable variation in the solution times for the various methods, but all methods succeed in locating an equilibrium, even from poor starting points.

Finally, in Chapter 8 we attempt to provide some perspective on the collection of results obtained and suggest some possible directions for future research. Overall, our analysis reveals a basic theoretical dilemma in solving general equilibrium problems by the sequential methods studied. One group of methods produces iterates that converge from any starting point, but the sequence may converge to a non-equilibrium point. Another group produces iterates that may fail to converge, but successfully converging sequences do attain an equilibrium. Nonetheless, the successful performance of these methods on actual models (both in our experiments and those of others) suggests that the kinds of general equilibrium models formulated in practice possess certain favorable computational properties that theoretical analysis has yet to discover.

1.4 A word on acronyms, notation, and internal references

This section provides a convenient listing of important acronyms and notation to be used in the following chapters. Both here and throughout, references to other parts of this document will denote chapter, section, and subsection. For instance, the reference "Section 3.2.1" denotes Subsection 1 of Section 2 of Chapter 3.

Throughout this document, vectors should be taken as column vectors unless explicitly transposed, as in z^T. The one exception to this is gradient vectors, for which we use the row convention. In particular, the Jacobian matrix of a vector-valued function $f(z)$ will be denoted by $\nabla f(z)$, for which the (i,j)-th element is $\left[\frac{\partial f_i(z)}{\partial z_j}\right]$, where $f_i(z)$ denotes the i-th component of $f(z)$.

We also will make use of index set notation for identifying sub-vectors and submatrices. For instance, if J is a set of indices, z_J refers to the sub-vector of components of a vector z whose indices are in the set J. For matrices, M_{JK} identifies the submatrix of a matrix M determined by the rows with indices in set J and the columns with indices in set K. We will use the notation $M_{J\bullet}$ to indicate a submatrix containing all of the columns of M and the rows with indices in set J. Similarly, $M_{\bullet K}$ selects all of the rows but only the columns in set K. The cardinality of a finite set J is indicated by $|J|$. Set subtraction (finite or infinite) will be indicated by $J \setminus K$, meaning the set of all elements of set J that are not also in set K.

1.4.1 Acronyms

CGE	Computable General Equilibrium problem (Section 2.4)
NLCP	NonLinear Complementarity Problem
NLP	NonLinear Program
QP	Quadratic Program
LP	Linear Program
CNLP	equivalent NLP formulation of an NLCP (Section 3.1)
Eq-NLCP	NLCP formulation of CGE (Section 3.2)
Eq-NLP	NLP formulation of CGE (Section 3.3)
DEPF	Differentiable Exact Penalty Function (Section 5.2.1)
Eq-DEPF	DEPF formulation of CGE (Section 5.2.1)
SLCP	Sequence of Linear Complementarity Problems (Section 4.1)
SQP	Sequence of Quadratic Programs (Section 4.2)
SLP	Sequence of Linear Programs (Section 4.4)
SLTZ	linearization method based on Slutzky matrices (Section 7.2.2)

1.4.2 Notation

$\mathbf{R}^m_+$	nonnegative orthant of m-dimensional Euclidean space
$\perp$	symbol denoting a complementarity condition
z_J	elements of vector z with indices in set J
M_{JK}	submatrix of matrix M; row indices in set J, column indices in set K
$\|J\|$	cardinality of a finite set J
$J \setminus K$	all elements of set J not also in set K
e	a vector of ones
e_i	the i-th unit vector
m	number of commodities/prices
n	number of production activities
p	m-vector of prices
y	n-vector of production activity levels
v	artificial variable
$d(p)$	m-vector of demands net of endowments (and nonlinear supplies)
x	m-vector of approximate demands
A	$m \times n$ activity analysis matrix
h	price normalization vector ($h^\mathsf{T} p = 1$)
$\hat{p}$	Lagrange multipliers for supply/demand balances
$\hat{y}$	Lagrange multipliers for nonprofitability conditions
$\hat{v}$	Lagrange multiplier for price normalization

2

THE COMPUTABLE GENERAL EQUILIBRIUM PROBLEM (CGE)

In this chapter we build up the computable general equilibrium problem (CGE) from its component parts. This entails describing the relevant mathematical properties of the demand and production structures, discussing the convention for normalizing prices, and finally delineating the mathematical conditions that define an equilibrium solution. Much of this material is based on the excellent presentations by Kehoe [Keh 82] and Mas-Colell [MasC 85].

2.1 Properties of demand

Let m denote the number of commodities and associated prices represented in the equilibrium problem. Let p denote an m-vector of prices and $d(p)$ an m-vector of final demands as a function of prices. For generality and notational convenience, $d(p)$ represents final demand net of any initial endowments or price-sensitive nonlinear supplies. Typically, $d(p)$ is undefined at the origin but defined for all strictly positive prices. If $d(p)$ is not defined on all of $\mathbf{R}^m_+ \setminus \{0\}$, then some further assumption about boundary behavior is necessary in order to ensure the existence of equilibrium prices in the domain of $d(p)$. To this end, define a set $\mathcal{U}$ contained in the boundary of $\mathbf{R}^m_+$ over which $d(p)$ is undefined. (Note: $0 \in \mathcal{U}$.) The domain of definition of $d(p)$ is denoted by $\mathcal{D} = \mathbf{R}^m_+ \setminus \mathcal{U}$.

We consider demand functions, $d(p)$, that satisfy the following properties on $\mathcal{D}$:

(D1) homogeneity of degree zero $\quad (d(\lambda p) = d(p),\ \forall \lambda > 0)$

(D2) Walras' law $\quad (p^{\mathsf{T}} d(p) = 0)$

(D3) boundedness from below $\quad (\exists\, \tau > 0 : d(p) \geq -\tau)$

(D4) twice-continuous differentiability

(D5) boundary behavior $\quad$ (for any $\bar{p} \in \mathcal{U} \setminus \{0\}$ and any sequence $p^i \longrightarrow \bar{p},\ p^i \in \mathcal{D}$ it must be that $\|d(p^i)\| \longrightarrow \infty$).

Assumption (D5) is a technical point important to issues of existence and setting up machinery to investigate regularity and uniqueness of equilibria. It is perhaps most easily understood by example. If the demand functions are derived from a Cobb-Douglas utility function, the demand for any commodity becomes infinite as its price approaches zero. In this case, the set $\mathcal{U}$ is the entire boundary of $\mathbf{R}^m_+$.

As to the other properties, homogeneity of degree zero (D1) follows from the general equilibrium nature of the problem (i.e., all prices are endogenous) and an implicit assumption that money is neutral and has no direct effect on economic decisions. Walras' law (D2) is a condition that all income is spent. This property is inherited by the aggregate demand

function when it is true for the demand functions of individuals. The implicit assumption is utility-maximizing behavior with non-satiation. Boundedness from below (D3) is the economically meaningful condition that initial endowments (and price-sensitive nonlinear supplies) are finite. Property (D4) is a restrictive assumption in theory, but not so restrictive in terms of practical computations. The motivations for this limitation were discussed in Section 1.1.

The above domain definitions and properties require special interpretation in the case of economies with primary and/or intermediate commodities used only in the production sector. A *primary* commodity is one for which aggregate demand is constant and equal to the negative of the aggregate endowment. (There is a potential ambiguity in defining the demand for a primary commodity when the price is zero. Any quantity in the range between the total endowment and the amount actually used by the production sector is an acceptable definition. It is convenient to treat the demand as constant, and, in the presence of a free disposal assumption to be made below, this convention can be used without loss of generality.) An *intermediate* commodity is one for which aggregate demand is identically zero. Aggregate demand for final commodities is unaffected by the prices of intermediate commodities. In the presence of primary commodities, it is essential that the set $\mathcal{U}$ include all price vectors in which only primary commodities have positive prices. This rules out the rather anomalous situation in which consumers have income but anything they might want to buy is free. If intermediate commodities exist, all definitions and properties above are to be interpreted in terms of demand functions defined for primary and final commodities only. Prices of intermediate commodities do not enter the demand functions and consequently are irrelevant to the definition of the domain $\mathcal{D}$.

The special model structures that arise in the case of intermediate and primary commodities will be explicitly relevant only in Chapter 6. The rest of our results and discussions apply independently of the commodity classifications. As explicitly carrying the partitioning of commodity balances and prices would only make the notation more cumbersome, we will use the more general and simple notation.

The homogeneity property and Walras' law imply some strong relationships between the first- and second-order derivatives of the demand functions. Homogeneity of degree zero (D1) implies via Euler's law that:

(H1) $\nabla d(p)p = 0.$

Further differentiating this identity implies:

(H2) $\nabla d_i(p) = -p^{\mathsf{T}}\nabla^2 d_i(p).$

Differentiating Walras' law (D2) implies:

(W1) $\nabla^{\mathsf{T}} d(p)p = -d(p).$

Further differentiating this identity implies:

(W2) $\sum_i p_i \nabla^2 d_i(p) = -\left[\nabla d(p) + \nabla^{\mathsf{T}} d(p)\right].$

An important implication of properties (H1) and (W1) is that $\nabla d(p)$ can be symmetric only

if $d(p)$ is identically zero. We are not aware of any previous use being made of properties (H2) and (W2). For our purposes, these relationships prove instrumental in establishing the equivalence of two solution methods in Section 4.3.

The properties itemized in this section are the only restrictions placed by economic theory on differentiable aggregate excess demand functions. This generality of form is the principal source of difficulty in solving general equilibrium problems.

2.2 Properties of production

Let n denote the number of linear production activities in the equilibrium problem (excluding disposal activities). Let y denote an n-vector of production activity levels and A the associated $m \times n$ activity analysis matrix. In this case, the net output from production is defined by the polyhedral cone $\{Ay : y \geq 0\}$. As a function of prices, the vector of excess profits obtained per unit of operation of the various production activities is given by $A^{\mathsf{T}}p$.

Unless a CGE model is specifically designed to address issues of indivisibility and/or increasing returns to scale in production activity, it is generally assumed that the set of aggregate production possibilities is closed and convex and admits inactivity. These conditions are clearly met in the case of linear production considered in this research. The other characteristic assumptions are the following:

(A1) free disposal of excess supply

(A2) output requires input $\quad (Ay \geq 0,\ y \geq 0 \implies Ay = 0).$

Any departures from free disposal property (A1) are readily represented by defining appropriate resource-consuming activities. Property (A2) is the specialization to linear production of the economically meaningful condition that no combination of activities can result in a net output (of anything) without a net input (of something). This property is mathematically equivalent to the following statement in terms of price conditions:

(A2′) prices yielding no profits $\quad (\exists\, p > 0 : A^{\mathsf{T}}p \leq 0).$

Economically, this means that positive prices always exist which eliminate all excess profits.

2.3 Price normalization

Because of the neutrality of money in the class of general equilibrium problems under consideration, an equilibrium solution determines only relative prices. This invariance to price scale is reflected in the homogeneity of degree zero of the final demand functions, property (D1), and in the fact that the *relative* profitabilities of linear production activities ($A^{\mathsf{T}}p$) are invariant to a positive scaling of all prices. A price normalization is a mathematical construct employed to remove this degree of freedom and determine a price scale (however arbitrary).

In principle, any nonnegative function of prices $h(p)$ which is homogeneous of degree one can be used to define a price normalization: $h(p) = constant$. It is somewhat traditional in

economic literature to employ a *numéraire* good, that is, a selected commodity whose price is arbitrarily fixed (at unity, say). Mas-Colell [MasC 85] makes general use of the unit ball: $\|p\| = 1$. Many others, including Kehoe [Keh 82], employ the unit simplex: $e^T p = 1$, where e is a vector of ones.

Given the inherent arbitrariness of the mode of price normalization, it makes sense to select a normalization which imparts desirable theoretical and/or computational properties to the equilibrium system. For present purposes we employ a general linear normalization: $h^T p = 1$, where $0 \neq h \geq 0$. Note that this includes the unit simplex and specifying a numéraire good as special cases, simply by setting $h = e$ or $h = e_i$, where e_i is the i-th unit vector and i is the index of the numéraire commodity. In Chapter 6 we build upon earlier work of the author [Sto 85] and of Eaves [Eav 87] in using the price normalization to computational advantage. In particular, we demonstrate that normalizing on a simplex of prices for final and primary commodities (in conjunction with including a matched artificial column to be discussed in the next chapter) is sufficient to guarantee the feasibility and solvability of subproblems derived by linearizing the final demand functions in the equilibrium problem.

2.4 General equilibrium conditions

Given the definition of demand and production components provided above, we can now delineate the mathematical conditions which characterize a general equilibrium solution. A general equilibrium is a set of nonnegative prices p and nonnegative production levels y which satisfy:

(E1)	$d(p)$	$\leq$	Ay	(demand cannot exceed supply)
(E2)	$A^T p$	$\leq$	0	(no positive excess profits)
(E3)	$h^T p$	$=$	1	(price normalization)
(E4)	$p^T d(p)$	$=$	$p^T Ay$	(excess supply has a zero price and a positive price means the market clears)
(E5)	$y^T A^T p$	$=$	0	(activities used have zero profit and no unprofitable activities are used).

Note that free disposal is reflected in (E4) and in the use of an inequality in (E1). Condition (E5) is in fact implied by (E4) and Walras' law (D2). Changing the price normalization in (E3) does not affect the essential characteristics of an equilibrium solution.

Conditions (E4) and (E5) are called complementarity conditions. Each inequality in (E1) and (E2) is associated with a matching price or activity variable, respectively. Complementarity of a given pair means that if the inequality is strict, the matching variable must be zero, and, conversely, if the variable is positive, the inequality must be satisfied as an equality. (This is equivalent to multiplying through each inequality by its matching variable and requiring the result to be an equality.) Conditions (E4) and (E5) are precisely the summing of these pairwise complementarity conditions for inequalities (E1) and (E2), respectively.

Since all inequalities have the same sense and all variables are nonnegative, equality holds for the sum if and only if all matching pairs are individually complementary.

While complementarity is an attribute serving to define an equilibrium solution, an interesting aspect of the general equilibrium problem is that any nonnegative feasible solution of (E1)–(E2) automatically satisfies the complementarity conditions (E4)–(E5). Because of Walras' law, the left-hand-side of (E4) is identically zero, making (E4) in effect the same as (E5). For the same reason, multiplying through inequality (E1) by any $p \geq 0$ implies $p^{\mathsf{T}}Ay \geq 0$. In turn, multiplying through inequality (E2) by any $y \geq 0$ implies $p^{\mathsf{T}}Ay \leq 0$. Hence, $p^{\mathsf{T}}Ay = 0$ and any nonnegative feasible solution of (E1)–(E2) satisfies (E4)–(E5).

The upshot of this result is that the general equilibrium problem is, in essence, a (nonlinear, nonconvex) feasibility problem. This characteristic will resurface a number of times in the discussions of later chapters. Nonetheless, the complementarity aspect remains central to one powerful and influential approach to formulating and solving the computable general equilibrium problem. It also distinguishes this formulation from specifications made purely in terms of simultaneous equations with no nonnegativity restrictions. Such specifications do not admit linear production and must presume that any solution will obtain positive prices. The complementarity formulation could be reduced to simultaneous equations given prior information as to the binding inequalities, which may not be limited to those for which the complementary prices and activities are positive at the equilibrium solution.

2.5 Existence of equilibrium solutions

The properties of demand and production delineated above are sufficient to ensure the existence of equilibrium solutions in the domain $\mathcal{D}$. Proofs are traditionally based on a fixed-point argument using either the Brouwer or the Kakutani fixed-point theorem (depending on the continuity of a map constructed on a compact, convex subset of the price space). The most appropriate proof for the present context is probably that of Kehoe [Keh 82], which explicitly deals with boundary behavior and assumption (D5), as well as with primary and intermediate commodities.

Note that the polyhedral region defined by conditions (E2) and (E3) is certainly closed and convex. To satisfy the assumptions for an existence proof, the price normalization is used to bound the region, as is easily done, for instance, by using the unit simplex ($h = e$). (It is also possible to normalize only the prices of final and primary commodities; see [Keh 82].) Once existence has been established as a theoretical matter, however, it is no longer necessary to use the same price normalization for computational purposes. Indeed, a normalization applied only to a subset of the prices may well prove advantageous, as we will show in Chapter 6. The important proviso here is that positive scalings of (at least one of) the equilibria existing on the unit simplex (say) be admitted by the alternative price normalization employed for computation. The only situation in which this would not be true is if all prices associated with positive components of $h \not> 0$ are in fact zero in all equilibrium solutions. The most likely example of this unfortunate circumstance would be choosing as the numéraire good a commodity which must have zero price at equilibrium.

Throughout the following discussions, it will be assumed that h has been chosen so as not to rule out the equilibrium solutions. This is not a restrictive assumption, since $h = e$ can always be used in the absence of enough insight into model structure to allow a better choice. Consequently, ***existence will not be an issue*** in any of the succeeding discussions of alternative mathematical formulations of the general equilibrium problem.

3

TWO EQUIVALENT FORMULATIONS OF CGE

In this chapter we develop two equivalent formulations of the computable general equilibrium problem (CGE). The first follows closely the statement of CGE in the previous chapter and comprises a specially structured nonlinear complementarity problem (NLCP). The second is a new development comprising a nonlinear program (NLP) derived from the NLCP by minimizing the nonnegative sum of the complementarity conditions over the same inequalities. Both formulations depart from more traditional formulations by including an artificial variable in the inequalities representing the supply/demand balances. This artificial variable proves to be helpful in two respects: (1) providing a simple characterization of an equilibrium solution, and (2) ensuring the solvability of subproblems generated by iterative methods for solving either formulation. Discussion of point (2) is deferred to Chapter 6.

First we briefly review the general form of an NLCP. This is partly to establish notation and partly to demonstrate some useful properties which later will be specialized to the formulation of CGE.

3.1 The nonlinear complementarity problem

The general nonlinear complementarity problem is defined as follows. Given a vector-valued function $f(z)$ defined for $z \geq 0$, find a solution that satisfies:

$$\text{(NLCP)} \qquad z \geq 0, \quad f(z) \geq 0 \quad \text{and} \quad z^{\mathsf{T}} f(z) = 0.$$

Any z satisfying the two nonnegativity conditions is called a *feasible* solution. For any such feasible solution it must be the case that $z_i f_i(z) \geq 0$ for all i and hence that $z^{\mathsf{T}} f(z) \geq 0$. A *complementary* solution is a feasible solution that satisfies the complementarity condition $z_i f_i(z) = 0$ for all i. Because of the nonnegativity of each term, this condition is equivalent to $z^{\mathsf{T}} f(z) = 0$.

It is well known that the general nonlinear complementarity problem may be equivalently stated as a nonlinear program:

$$\text{(CNLP)} \qquad \text{minimize} \quad z^{\mathsf{T}} f(z) \quad \text{subject to} \quad z \geq 0, \quad f(z) \geq 0.$$

Note that the objective function of (CNLP) is bounded below by zero on the feasible region. Given the existence of a solution of (NLCP), it follows that global minimizers of (CNLP) are complementary solutions of (NLCP) and vice versa.

The special case of an affine $f(z) = Mz + q$ produces a linear complementarity problem, for which the corresponding CNLP form is a quadratic program with Hessian $M + M^{\mathsf{T}}$.

Mathematical programs of the CNLP form have been called *composite* programs by Cottle (e.g., in [Cot 64]), whose original studies of the form pertained to the optimality conditions

for convex quadratic programs. More generally, the (first-order) optimality conditions for any NLP can be represented by an NLCP, and this establishes an important connection between nonlinear complementarity and nonlinear programming. It is essential, however, not to confuse this connection for a special class of NLCPs with the constructed equivalence between problems (NLCP) and (CNLP) above, which applies regardless of the origin and specific form of (NLCP).

3.2 CGE as a nonlinear complementarity problem (Eq-NLCP)

We now transform the statement of equilibrium conditions for CGE in Section 2.4 into a highly structured NLCP. We begin by introducing an artificial variable v into the supply/demand balance inequalities (E1). The associated artificial column has the appearance of an extra linear supply activity in which, for reasons which will become apparent, the output coefficients are defined by the price normalization vector h. After some stylistic reformatting, the CGE problem can be equivalently formulated as the following nonlinear complementarity problem:

Eq-NLCP

Find a complementary solution of:

$$
\begin{array}{lrcl}
\text{(NLCP1)} & -d(p) + Ay + hv \geq 0 & \perp & p \geq 0 \\
\text{(NLCP2)} & -A^{\mathsf{T}}p \geq 0 & \perp & y \geq 0 \\
\text{(NLCP3)} & -h^{\mathsf{T}}p = -1 & \perp & v \\
\text{(NLCP4)} & p \ , \ y \geq 0 &&
\end{array}
$$

Here we place the shorthand notation "$\perp \ p \geq 0$" to the right of an inequality in order to indicate that each component of the nonnegative vector p is to be pairwise complementary with the (slack variable of the) associated inequality.

We have innocuously changed the sign of the price normalization constraint so as to produce a skew-symmetric structure in the two linear blocks of the system of inequalities. The addition of variable v makes the system square, and its associated inequality is actually the equality price normalization constraint. Consequently, the variable v is not sign-constrained, and the complementarity condition is always satisfied. This deviates somewhat from the standard format of a complementarity problem, but it seems best to avoid the notational tedium of splitting the price normalization into two inequalities and v into its positive and negative parts.

The complementarity conditions for Eq-NLCP reduce to a very simple form. To obtain the appropriate specialization of "$z^{\mathsf{T}}f(z)$," take any feasible solution and multiply through

inequalities (NLCP1) by $p \geq 0$ and inequalities (NLCP2) by $y \geq 0$. Add the two resulting scalar inequalities to obtain:

$$-p^{\mathsf{T}}d(p) + p^{\mathsf{T}}Ay + p^{\mathsf{T}}hv - y^{\mathsf{T}}A^{\mathsf{T}}p \geq 0.$$

Now note that the first term vanishes because of Walras' law, (D2) of Section 2.1. The second and fourth terms cancel, and $p^{\mathsf{T}}h = 1$ by (NLCP3). What is left is simply $v \geq 0$. Hence, any feasible solution of Eq-NLCP necessarily has $v \geq 0$, and the complementary solutions of Eq-NLCP are those feasible solutions with $v = 0$.

The correspondence of Eq-NLCP to the equilibrium conditions for CGE in Section 2.4 is immediate. Inequalities (NLCP2) are precisely CGE conditions (E2), and complementarity with y is exactly (E5). (NLCP3) is just (E3) with a sign reversal. Inequalities (NLCP1) correspond to (E1), with the addition of an artificial term hv. Complementarity with p corresponds logically to (E4). Despite initial appearances, the presence of hv does not disturb the equivalence at equilibrium. Since any feasible solution of Eq-NLCP has $v \geq 0$, this solution may not satisfy (E1). However, a complementary solution has $v = 0$ and thus directly satisfies both (E1) and (E4).

It follows then that equilibrium solutions of CGE can be computed by solving problem Eq-NLCP. It is important to keep in mind the implication of properties (H1) and (W1) of Section 2.1 that $d(p)$ can *never* be the gradient of any scalar-valued function. Consequently, even though the linear parts of the problem conform to a skew-symmetric structure, Eq-NLCP does not represent the first-order conditions for optimality of any nonlinear program.

3.3 CGE as a nonlinear program (Eq-NLP)

Constructing the CNLP nonlinear program corresponding to problem Eq-NLCP is particularly easy since the complementarity conditions reduce to $v = 0$. Using this fact, we may define the following nonlinear programming formulation of the computable general equilibrium problem:

Eq-NLP

	minimize subject to									
						v				
(NLP1)		$-d(p)$	$+$	Ay	$+$	hv	$\geq$	0	$\perp$	$\hat{p} \geq 0$
(NLP2)		$-A^{\mathsf{T}}p$					$\geq$	0	$\perp$	$\hat{y} \geq 0$
(NLP3)		$-h^{\mathsf{T}}p$					$=$	-1	$\perp$	$\hat{v}$
(NLP4)		p	,	y			$\geq$	0		

Note that the inequalities of Eq-NLP are identical to those of Eq-NLCP. At a Karush-Kuhn-Tucker point of Eq-NLP, each inequality is complementary with an associated Lagrange

multiplier, and the collection of such multipliers is suggestively denoted by $(\hat{p}, \hat{y}, \hat{v})$. By the same argument as made above, $v \geq 0$ for any feasible solution of Eq-NLP, and global minima are those feasible solutions that attain $v = 0$.

Since global minimizers of Eq-NLP are equivalent to complementary solutions of Eq-NLCP, the correspondence to equilibrium solutions of CGE is the same as shown above. The form of problem Eq-NLP makes explicit the observation of Section 2.4 that the equilibrium conditions of CGE in essence define a feasibility problem. Eq-NLP is precisely seeking a feasible solution of conditions (E1)-(E3) by minimizing the value of an artificial variable introduced into (E1).

3.3.1 Stationary points of Eq-NLP

Aggregate demand functions are not in general convex, and finding a global minimum of a nonconvex program can be an extremely difficult undertaking in practice. In the case of problem Eq-NLP, we at least have the prior knowledge that global minima have $v = 0$, so stationary points found by any applied solution algorithm can be readily and definitively checked for optimality. In addition, the special structure of Eq-NLP and the properties of $d(p)$ impart some interesting properties to stationary points of Eq-NLP which suggest that nonoptimal stationary points may be rare or at least avoidable. While we have not been able to obtain any conclusive results in this regard, it is worthwhile discussing some partial results as a starting point (and hopefully motivation) for future research into this issue.

Let $(\bar{p}, \bar{y}, \bar{v})$ be a Karush-Kuhn-Tucker (KKT) point of Eq-NLP, with $(\hat{p}, \hat{y}, \hat{v})$ the associated Lagrange multipliers. For the present we consider only the first-order conditions, which entail the following parallel sets of inequalities and complementarity relationships:

$$
\begin{array}{llll|llll}
\text{(P1)} & -d(\bar{p}) + A\bar{y} + h\bar{v} \geq 0 & \perp \hat{p} \geq 0 & & \nabla^{\mathsf{T}} d(\bar{p})\hat{p} + A\hat{y} + h\hat{v} \geq 0 & \perp \bar{p} \geq 0 & \text{(M1)} \\
\text{(P2)} & -A^{\mathsf{T}}\bar{p} \geq 0 & \perp \hat{y} \geq 0 & & -A^{\mathsf{T}}\hat{p} \geq 0 & \perp \bar{y} \geq 0 & \text{(M2)} \\
\text{(P3)} & -h^{\mathsf{T}}\bar{p} = -1 & \perp \hat{v} & & -h^{\mathsf{T}}\hat{p} = -1 & \perp \bar{v} & \text{(M3)} \\
\text{(P4)} & \bar{p}\ ,\ \bar{y} \geq 0 & & & \hat{p}\ ,\ \hat{y} \geq 0 & & \text{(M4)}
\end{array}
$$

Equivalent statements of (P1):

$$
\begin{array}{ll}
\text{(P1W)} & \nabla^{\mathsf{T}} d(\bar{p})\bar{p} + A\bar{y} + h\bar{v} \geq 0 \perp \hat{p} \geq 0 \\
\text{(P1H)} & -\nabla d(\bar{p})\bar{p} + A\bar{y} + h\bar{v} \geq d(\bar{p}) \perp \hat{p} \geq 0
\end{array}
$$

The equivalence of conditions (P1W) and (P1H) to (P1) follows from properties (W1) and (H1) of Section 2.1. We will refer to the conditions on the left as the primal conditions (hence the labeling with "P") and to those on the right as the multiplier conditions (hence the "M"). Because of the nonconvexity of Eq-NLP, the above conditions need not be sufficient for optimality of the solution.

We now detail a number of observations about the KKT system for Eq-NLP.

Observation 1. Putting (P1W) in place of (P1) demonstrates that the primal inequalities are identical in form to the multiplier inequalities at a KKT point. This implies in particular that the primal feasible solution $(\bar{p}, \bar{y}, \bar{v})$ is also feasible with respect to the multiplier inequalities (M1)–(M4). It may not be self-complementary, however, in which case $\bar{v} > 0$. As we have already shown, a primal feasible solution is globally minimal (i.e., complementary) if and only if $\bar{v} = 0$.

Observation 2. Putting (P1H) in place of (P1) gives primal and multiplier conditions which are precisely the complementary slackness conditions for the linear program of minimizing v subject to "generic" versions of (P1H) and (P2)–(P4). This linear program will resurface in Section 4.4.

Observation 3. We may intuitively expect that, at any KKT point, $\hat{v} = 0$. The reasoning is that $\hat{v}$ is the Lagrange multiplier for price normalization (NLP3) in Eq-NLP, and neither (NLP1) nor (NLP2) is affected by changing the price scale. This intuition can in fact be validated as follows. By complementarity of $\bar{p}$ with (M1), we have

$$0 = \bar{p}^{\mathsf{T}} \nabla^{\mathsf{T}} d(\bar{p}) \hat{p} + \bar{p}^{\mathsf{T}} A \hat{y} + \bar{p}^{\mathsf{T}} h \hat{v} = 0 + 0 + \hat{v}.$$

Here, the first term vanishes because of (H1) of Section 2.1, the second term vanishes because $\hat{y}$ is complementary with (P2), and the third term reduces to $\hat{v}$ because of (P3).

Observation 4. By a similar argument using complementarity of $\hat{p}$ with (P1), we obtain $\bar{v} = \hat{p}^{\mathsf{T}} d(\bar{p}) = -\bar{p}^{\mathsf{T}} \nabla d(\bar{p}) \hat{p}$, where the second equality results from transposition and (W1) of Section 2.1. Since $\bar{v} > 0$ at any non-optimal (non-equilibrium) solution, we may derive the economic interpretation that the consumption bundle $d(\bar{p})$ would not be affordable at prices $\hat{p}$.

Observation 5. By multiplying through (M1) and (M2) by $\hat{p} \geq 0$ and $\hat{y} \geq 0$, respectively, and then adding the result, we obtain $\hat{p}^{\mathsf{T}} \nabla d(\bar{p}) \hat{p} \geq 0$. Adding this inequality to the expression obtained for $\bar{v}$ in Observation 4, and again making use of (H1) from Section 2.1, we can conclude that $\bar{v} \leq (\hat{p} - \bar{p})^{\mathsf{T}} \nabla d(\bar{p})(\hat{p} - \bar{p})$. In particular this implies that, if by some chance $\nabla d(\bar{p})$ is negative semidefinite on the linear subspace orthogonal to h, then $\bar{v}$ must be zero.

We will again have use for these KKT conditions in Chapter 5, wherein we demonstrate that the multipliers from a nonoptimal stationary point of Eq-NLP can be used to construct a descent direction in a penalty function problem that is equivalent to Eq-NLP.

4

SOLUTION METHODS FOR Eq-NLCP AND Eq-NLP

In this chapter we examine the application of five general-purpose methods to solving the nonlinear complementarity problem (Eq-NLCP) or nonlinear program (Eq-NLP) formulations of the computable general equilibrium problem. The common feature of these methods is solving a sequence of subproblems defined by linearizing nonlinear constraints, i.e., the aggregate demand functions in the equilibrium problem. The methods may be distinguished by the type of linearization employed and the kind of solution sought relative to the linearized constraints. The first four methods employ a first-order Taylor's expansion and then solve an appropriately defined linear complementarity problem (LCP), quadratic program (QP), linear program (LP), or projected Lagrangian problem, respectively. The fifth class of methods employs any of a number of alternative linearizations and then solves either an LCP or a QP. We defer the discussion of convergence of these methods to the next chapter.

Our list of methods by no means includes all possible optimization methods for solving problem Eq-NLP. In particular we omit sequential unconstrained methods based on penalty or barrier functions (although such functions may appear as merit functions guiding one of the other methods considered). For the most part, this reflects a bias towards exploiting as much as possible the largely linear nature of the CGE problem under study.

4.1 A sequence of linear complementarity problems (SLCP)

The notion of solving a nonlinear complementarity problem by generating and solving a sequence of linear complementarity problems is a natural analog of well-studied methods for solving systems of nonlinear equations by sequences of linear equations. In particular, the analog of Newton's method is based on employing a first-order Taylor's expansion as the mode of linearization. Important contributions to the extension of Newton's method to NLCPs are found in the algorithm of Eaves [Eav 78] and in the work of Josephy and Robinson on generalized equations, which is comprehensively surveyed in [Rob 82]. The NLCP is an important special case of the generalized equation, and Josephy in [Jo-N 79] demonstrates that the application of Newton's method to the generalized equations corresponding to an NLCP results in solving a sequence of LCPs.

Mathiesen independently elaborated the SLCP method as a means of solving economic equilibrium problems. The best presentation of his work from a theoretical perspective is in [Mat 87]. Particularly in collaboration with Rutherford, Mathiesen has conducted extensive empirical investigations into the behavior of SLCP as applied to both partial and general equilibrium problems. Almost all of the computational results attest to the general robustness and efficiency of the SLCP method. The most informative discussions of these experiments are unfortunately in unpublished form, [MR 83] and [Rut 86].

SLCP is applied to problem Eq-NLCP as follows. Let an initial linearization point be given as p^0, which must be in the domain of $d(p)$ but need not be feasible with respect to the profitability conditions (NLCP2) in Section 3.2. Thereafter, any iteration $k+1$ begins with a first-order Taylor's approximation of $d(p)$ at the point p^k:

$$d(p) \approx d^k(p) \equiv d(p^k) + \nabla d(p^k)(p - p^k) = d(p^k) + \nabla d(p^k)p,$$

where the simplification on the right results from (H1) of Section 2.1. We then define and solve the linearized subproblem denoted by LCP(p^k):

$$\boxed{\text{LCP}(p^k)}$$

Find a complementary solution of:

$$\begin{array}{lrcl} \text{(LCP1)} & -\nabla d(p^k)p + Ay + hv & \geq & d(p^k) \quad \perp \quad p \geq 0 \\ \text{(LCP2)} & -A^{\mathsf{T}}p & \geq & 0 \quad \perp \quad y \geq 0 \\ \text{(LCP3)} & -h^{\mathsf{T}}p & = & -1 \quad \perp \quad v \\ \text{(LCP4)} & p \ , \ y & \geq & 0 \end{array}$$

Let $(\bar{p}, \bar{y}, \bar{v})$ solve LCP(p^k). If $\bar{p}$ is sufficiently close to p^k, then we have an approximate complementary solution of Eq-NLCP. If not, we define a new linearization point by:

$$p^{k+1} = p^k + \lambda(\bar{p} - p^k), \quad \text{where} \quad 0 < \lambda \leq 1.$$

Set $k = k + 1$ and repeat.

As in most iterative algorithms, the choice of step length λ is a matter of art. At the least, λ must be chosen so as to ensure that p^{k+1} remains in the domain of $d(p)$. We will have only a few more remarks about the issue of step length in the next chapter on convergence.

An important consideration in the definition and execution of SLCP is the feasibility and solvability of the LCPs encountered. Mathiesen uses the numéraire method of price normalization and Lemke's almost-complementary pivoting method [Lem 68] for solving each LCP. He resorts to choosing a different numéraire commodity whenever a subproblem is not solved. As we shall demonstrate in Chapter 6, however, the normalization vector h can be chosen in such a way as to guarantee that subproblems constructed as above have a feasible complementary solution which moreover can be successfully computed by Lemke's method.

4.1.1 Basic solutions, uniqueness, and regularity

Since Lemke's method proceeds by examining only basic solutions of an LCP, it is important to ascertain whether an equilibrium solution can in fact occur at an extreme point (or vertex) of the linearized contraints. Mathiesen correctly recognized that, in the absence of a price normalization, the equilibrium conditions do not admit a basic solution of the

inequalities. (We will review this observation below.) In practice Mathiesen avoids the singularity by deleting the price column and the supply/demand inequality corresponding to the numéraire commodity. We will show that adding the explicit price normalization equality (LCP3) and the artificial variable v accomplishes the same purpose. In fact, the two approaches are equivalent in this respect for $h = e_i$. There is, however, a minimum rank condition required, at least at an equilibrium price.

Since any equilibrium solution corresponds to a complementary solution of Eq-NLCP, the only question is whether it can be reduced to a basic complementary solution. That is, given an equilibrium price vector $\bar{p}$, can we construct a basic complementary solution of LCP($\bar{p}$)? To this end, let β^+ be an index set for all positive components of $\bar{p}$ and also for the corresponding rows of the matrix A. Let α index all those columns of A such that $\bar{p}_{\beta^+}^{\mathsf{T}} A_{\beta^+\alpha} = 0$. Accordingly, in any complementary solution, $\bar{y}_j = 0$ for all $j \notin \alpha$. We further partition the set α into sets α^+ and $\alpha^0 = \alpha \setminus \alpha^+$. This partition can always be constructed so as to satisfy the property that α^+ is a maximal linearly independent set of columns such that there exists a $(\bar{y}_{\alpha^+}, \bar{y}_{\alpha^0})$ satisfying:

$$\begin{aligned} A_{\beta^+\alpha^+}\bar{y}_{\alpha^+} &= d_{\beta^+}(\bar{p}) \\ A_{\beta^0\alpha^+}\bar{y}_{\alpha^+} &= d_{\beta^0}(\bar{p}) \\ A_{\beta^-\alpha^+}\bar{y}_{\alpha^+} &> d_{\beta^-}(\bar{p}) \\ \bar{y}_{\alpha^+} &> 0 \\ \bar{y}_{\alpha^0} &= 0\,. \end{aligned}$$

Here, the index set β^+ is given, while the index sets β^0 and β^- are implicitly defined to correspond to weakly binding ($\bar{p}_{\beta^0} = 0$) and strictly slack inequalities in the above solution.

We now construct a partitioned matrix which will be the focus of the ensuing discussion. For brevity we let $\nabla = \nabla d(\bar{p})$.

$$\left[\begin{array}{ccc|cc} -\nabla_{\beta^+\beta^+} & A_{\beta^+\alpha^+} & h_{\beta^+} & -\nabla_{\beta^+\beta^0} & A_{\beta^+\alpha^0} \\ -(A_{\beta^+\alpha^+})^{\mathsf{T}} & & & -(A_{\beta^0\alpha^+})^{\mathsf{T}} & \\ -h_{\beta^+}^{\mathsf{T}} & & & -h_{\beta^0}^{\mathsf{T}} & \\ \hline -\nabla_{\beta^0\beta^+} & A_{\beta^0\alpha^+} & h_{\beta^0} & -\nabla_{\beta^0\beta^0} & A_{\beta^0\alpha^0} \\ -(A_{\beta^+\alpha^0})^{\mathsf{T}} & & & -(A_{\beta^0\alpha^0})^{\mathsf{T}} & \end{array}\right] \equiv \left[\begin{array}{cc} M_{JJ} & M_{JK} \\ M_{KJ} & M_{KK} \end{array}\right]$$

The abbreviated notation in terms of "M" corresponds to that of Mangasarian [Mang 80], wherein J indexes binding inequalities with complementary variables that are positive and K indexes binding inequalities with complementary variables that are zero. Since the price normalization is an equality and variable v is unconstrained in sign (and therefore necessarily basic), it is appropriate to include them in set J — even though $\bar{v} = 0$ at an equilibrium solution.

In building a basis for this solution, the J columns must always be included while the K columns need not be considered at all. Indeed, a satisfactory and sparser set of basis columns

corresponding to the K rows would be the columns for the (degenerate) slack variables of the K rows, specifically the β^0 rows of (LCP1) and the α^0 rows of (LCP2). In addition, nondegenerate slacks are in the basis for all inequalities that are strict. Consequently, the ability to construct a basic solution turns on the nonsingularity of M_{JJ}. We must then address the conditions under which this is true.

Let M_{++} denote the leading principal submatrix of M_{JJ} obtained by deleting the row and column containing h_{β^+}. The submatrix M_{++} is singular because $\nabla_{\beta^+\beta^+}\bar{p}_{\beta^+} = 0$ and $(A_{\beta^+\alpha^+})^{\mathsf{T}}\bar{p}_{\beta^+} = 0$ at an equilibrium solution. Also, by virtue of property (W1) from Section 2.1, $(\nabla_{\beta^+\beta^+})^{\mathsf{T}}\bar{p}_{\beta^+} + A_{\beta^+\alpha^+}\bar{y}_{\alpha^+} = 0$. Hence, $(\bar{p}_{\beta^+}^{\mathsf{T}}\ \bar{y}_{\alpha^+}^{\mathsf{T}})M_{++} = 0$. Thus, since $h_{\beta^+}^{\mathsf{T}}\bar{p}_{\beta^+} = 1$, we can conclude:

$$M_{++}z + \begin{pmatrix} h_{\beta^+} \\ 0 \end{pmatrix} w = 0 \implies 0 + w = 0,$$

that is, that the artificial column is linearly independent of the columns of M_{++}. Nonsingularity of M_{JJ} then depends on the rank of M_{++}. If the rank of M_{++} is $|\beta^+| + |\alpha^+| - 1$ (where $|\cdot|$ denotes cardinality), then the null space of M_{++} is spanned by the single vector $(\bar{p}_{\beta^+}, 0)$. But since $h_{\beta^+}^{\mathsf{T}}\bar{p}_{\beta^+} = 1$, we can conclude from this and the preceeding implication that bordering M_{++} with the price normalization and artificial column serves to make M_{JJ} nonsingular. If M_{++} has more than one degree of rank deficiency, then naturally we cannot expect the addition of one row and column to produce a nonsingular matrix. Thus, the necessary and sufficient condition for existence of a *basic* equilibrium solution is that the associated submatrix M_{++} have rank $|\beta^+| + |\alpha^+| - 1$. This argument was inspired by a similar argument of Mas-Colell [MasC 85] for a pure exchange economy.

The above conclusion holds for any normalization such that $0 \neq h_{\beta^+} \geq 0$. In particular, it applies for $h = e_i$, for any $i \in \beta^+$. By a simple expansion of determinants argument, it is clear that M_{JJ} is nonsingular in this case if and only if the submatrix of M_{++} obtained by deleting the row and column corresponding to index i is nonsingular. Hence, Mathiesen's procedure of deleting the column and row corresponding to the numéraire price and commodity balance is equivalent to adding an explicit price normalization and matching artificial column based on $h = e_i$, where i is the index of the numéraire commodity.

We are thus left with the reassurance that a solution procedure which examines only basic solutions can find any equilibrium solution that satisfies the indicated rank condition on the submatrix M_{++}. Note that this condition depends on both the Jacobian of the demand functions and the activity analysis matrix.

We can also use the above partitioned matrix to examine a solution for local uniqueness and regularity. Mangasarian in [Mang 80] derives a collection of necessary/sufficient and sufficient conditions for local uniqueness of a solution of an LCP. These conditions are in turn sufficient for local uniqueness of the solution of an NLCP for which the LCP is a linearization at the solution point. Mangasarian employs the CNLP form of the complementarity problem and assumes twice-continuous differentiability so as to utilize the second-order sufficient conditions for optimality. Kyparisis in [Kyp 86] obtains the same results assuming only once-continuous differentiability by using the theory of generalized equations. Robinson in [Rob 80] establishes related conditions (and definitions) for the regularity of solutions

to complementarity problems as important special cases of regular solutions of generalized equations.

Mangasarian's necessary and sufficient conditions for local uniqueness of a solution of an LCP solution encompass both the general case and the particular case of M_{JJ} nonsingular. The nonsingular case is the most interesting in our context, both because we examine solution algorithms that find basic solutions and because there is then an immediate connection to Robinson's conditions for regularity. Both issues turn on the properties of the Schur complement of M_{JJ} in the above partitioned matrix, which we denote by $S \equiv M_{KK} - M_{KJ}M_{JJ}^{-1}M_{JK}$. A solution is locally unique if and only if $z = 0$ is the only solution of the LCP:

$$\text{find} \quad z \geq 0 \quad \text{such that} \quad Sz \geq 0 \quad \text{and} \quad z^{\mathsf{T}}Sz = 0.$$

A solution is regular if and only if all of the principal minors of S are positive. Clearly, regularity then implies local uniqueness.

In the further special case of nondegenerate (strict complementary slack) solutions, we have $K = \emptyset$. In this case regularity is equivalent to nonsingularity of M_{JJ}. Such solutions are what Kehoe [Keh 82] and Mas-Colell [MasC 85] refer to as regular and properly regular equilibria, respectively. Their focus on the nondegenerate case is necessitated by the extensive use of differential topology and genericity analysis based on arbitrary perturbations of problem data. There is some reassurance in their findings that regularity is a generic result in the space of production economies, meaning that almost all economies have only regular equilibria. Regularity is significant for computational purposes not only in guaranteeing the existence of basic equilibrium solutions but also in providing conditions for convergence of solution methods to be discussed in the next chapter.

4.2 A sequence of quadratic programs (SQP)

An important class of methods for solving general nonlinear programs proceeds by solving a sequence of quadratic programs, derived by linearizing the nonlinear constraints (via Taylor's expansion) and formulating a second-order approximation of the associated Lagrangian function. A reasonably self-contained discussion of various manifestations of the basic SQP method can be found in [GMW 81]. Different versions may be characterized by the nature of the Lagrangian approximation, whether or not the linearized constraints are perturbed in any way, and by the nature of the merit function and line search employed to promote convergence from any starting point. Typically, the QP subproblem is expressed in terms of a search direction relative to the linearization point.

We will not deal with constraint perturbations and defer the convergence issues to later chapters. With respect to problem Eq-NLP, the SQP process is very similar to applying SLCP to Eq-NLCP. Specifically, p^0 must be given and iteration $k+1$ begins by linearizing $d(p)$ at p^k. A second-order approximation of the Lagrangian function must be defined, noting that for problem Eq-NLP the only source of curvature is the demand functions $d(p)$ in inequalities (NLP1). We leave this approximation non-specific for the moment and denote the estimated Hessian matrix by $H(p^k, \hat{p}^k)$, thus indicating its dependence on both

the linearization point and an estimate of the Lagrange multipliers, $\hat{p}^k$. We then define and solve the linearized subproblem denoted by QP(p^k):

$$\boxed{\text{QP}(p^k)}$$

$$\begin{array}{lll} & \text{minimize} & \tfrac{1}{2}(p-p^k)^{\mathsf{T}} H(p^k,\hat{p}^k)(p-p^k) \;+\; v \\ & \text{subject to} & \\ \text{(QP1)} & & -\nabla d(p^k)p \;+\; Ay \;+\; hv \;\geq\; d(p^k) \quad \perp \quad \hat{p} \geq 0 \\ \text{(QP2)} & & -A^{\mathsf{T}}p \;\geq\; 0 \quad \perp \quad \hat{y} \geq 0 \\ \text{(QP3)} & & -h^{\mathsf{T}}p \;=\; -1 \quad \perp \quad \hat{v} \\ \text{(QP4)} & & p \;,\; y \;\geq\; 0 \end{array}$$

Let $(\bar{p},\bar{y},\bar{v})$ solve QP(p^k), with $(\hat{p},\hat{y},\hat{v})$ the associated Lagrange multipliers. If $\bar{p}$ and $\hat{p}$ are sufficiently close to p^k and $\hat{p}^k$, then we have an approximate local minimum for Eq-NLP. If not, we define a new linearization point by:

$$p^{k+1} = p^k + \lambda(\bar{p} - p^k), \quad \text{where} \quad 0 < \lambda \leq 1.$$

Typically, the Lagrange multiplier estimate is updated by $\hat{p}^{k+1} = \hat{p}$. Set $k = k+1$, and repeat.

Note that problem QP(p^k) has the same inequalities as LCP(p^k), so it inherits the same properties of feasibility and non-singularity as discussed above. In many practical implementations, the Hessian estimate $H(p^k,\hat{p}^k)$ is constructed so as to be positive semidefinite. If this is not enforced, as in the next subsection, it is desirable at least that the subproblem be constructed so that the objective function is bounded below on the feasible region.

4.3 SLCP implements SQP

In the original SQP method of Wilson [Wil 63], the QP objective function is derived from the second-order Taylor's expansion of the Lagrangian function at the linearization point p^k. The Lagrangian is constructed from the exact Hessians of the objective and constraint functions using the QP multipliers $\hat{p}^k$ as Lagrange multiplier estimates. The method was intended for use on convex problems, for which the resulting QPs would also be convex.

Even in the absence of convexity, we can specialize Wilson's formulation to problem Eq-NLP, yielding the following objective function for QP(p^k):

$$\text{minimize} \quad \tfrac{1}{2}(p-p^k)^{\mathsf{T}}\left[\sum_i \hat{p}_i^k \nabla^2 d_i(p^k)\right](p-p^k) + v.$$

If by design or chance $\hat{p}^k = p^k$, the Hessians can be eliminated using (W2) of Section 2.1 to produce:

$$\text{minimize} \quad -\tfrac{1}{2}(p-p^k)^{\mathsf{T}}\left[\nabla d(p^k) + \nabla^{\mathsf{T}} d(p^k)\right](p-p^k) + v.$$

We can then apply (H1) and (W1) from Section 2.1 to obtain:

$$\text{minimize} \quad -\tfrac{1}{2}p^{\mathsf{T}}\left[\nabla d(p^k) + \nabla^{\mathsf{T}} d(p^k)\right]p - p^{\mathsf{T}} d(p^k) + v,$$

which in non-symmetric form is:

$$\text{minimize} \quad -p^{\mathsf{T}}\nabla d(p^k)p - p^{\mathsf{T}} d(p^k) + v.$$

The resulting QP is thus precisely the composite quadratic program equivalent to LCP(p^k). The objective function is a statement of the complementarity conditions and consequently is bounded below by zero on the feasible region. Moreover, the global minima of the composite QP are equivalent to the complementary solutions of the LCP.

We can now provide an SQP interpretation of SLCP iterations. Suppose that at iteration 1 we take $\hat{p}^0 = p^0$. (Given the underlying complementarity nature of the equilibrium problem, this would seem to be a better estimate than most.) With this initialization, QP(p^0) and LCP(p^0) are identical problems. Given this equivalence, a complementary solution $(\bar{p}, \bar{y}, \bar{v})$ of LCP(p^0) is indeed a global minimizer of QP(p^0). Hence, if a unit step length is applied (as in Wilson's algorithm), for the next iteration we can take $p^1 = \hat{p}^1 = \bar{p}$. So again, QP($p^1$) and LCP($p^1$) are identical. By the obvious induction, this equivalence continues so long as a unit step length is taken. Thus, SLCP (with unit step length) is precisely implementing Wilson's SQP method on a nonconvex problem. The saving feature is that solving the subproblem as a complementarity problem guarantees finding a global minimum for the (composite) QP.

If a unit step length is not taken, the equivalence can be maintained by applying the same update formulas to the primal iterates p^k and the Lagrange multipliers $\hat{p}^k$. Such simultaneous updating is a feature of a different version of SQP implemented by Gill, Murray, Saunders and Wright [GMSW 86].

4.3.1 A more general equivalence result

We derived the equivalence result just discussed using the special properties of the demand functions in a general equilibrium problem. As it turns out, however, this result is actually a special case of an equivalence applying to any NLCP. Our discovery of the more general result was kindled by Mangasarian's utilization in [Mang 80] of the second-order sufficient optimality conditions for the equivalent CNLP statement of the general NLCP (defined in Section 3.1).

Consider applying Wilson's SQP method to problem CNLP. Let $\tilde{z}$ be a linearization point leading to the following linearized constraints:

$$z \geq 0 \quad \text{and} \quad \nabla f(\tilde{z})z \geq -f(\tilde{z}) + \nabla f(\tilde{z})\tilde{z}.$$

The QP objective function is formed from the gradient of the CNLP objective and an estimated Hessian of the Lagrangian function based on estimated Lagrange multipliers, $\tilde{z}$.

The gradient of the CNLP objective is:

$$\left[f(\bar{z}) + \nabla^{\mathsf{T}} f(\bar{z})\bar{z}\right]^{\mathsf{T}}.$$

The estimated Hessian of the Lagrangian is:

$$\nabla f(\bar{z}) + \nabla^{\mathsf{T}} f(\bar{z}) + \sum_i (\bar{z}_i - \hat{z}_i)\nabla^2 f_i(\bar{z}).$$

Again, if by design or chance $\hat{z} = \bar{z}$, the summation term drops from the estimated Hessian. We can then compose and simplify the QP objective function. The contribution from the gradient term is:

$$\left[f(\bar{z}) + \nabla^{\mathsf{T}} f(\bar{z})\bar{z}\right]^{\mathsf{T}}(z - \bar{z}) = z^{\mathsf{T}} f(\bar{z}) + \bar{z}^{\mathsf{T}}\nabla f(\bar{z})z + constant.$$

The contribution from the Hessian term is:

$$\tfrac{1}{2}(z - \bar{z})^{\mathsf{T}}\left[\nabla f(\bar{z}) + \nabla^{\mathsf{T}} f(\bar{z})\right](z - \bar{z}) = z^{\mathsf{T}}\nabla f(\bar{z})(z - \bar{z}) - \bar{z}^{\mathsf{T}}\nabla f(\bar{z})z + constant.$$

Collecting non-constant terms, we are left with

$$\text{minimize} \quad z^{\mathsf{T}}[\nabla f(\bar{z})z + f(\bar{z}) - \nabla f(\bar{z})\bar{z}].$$

Relating this to the linearized constraints, we see once again the composite quadratic program corresponding to the LCP derived from linearizing the NLCP at $\bar{z}$. The inductive argument given above then applies to the general case as well.

We summarize the equivalence result just obtained. Applying the SLCP method (with unit steps) to the general NLCP problem precisely implements Wilson's SQP method on the equivalent CNLP problem. Applying Wilson's SQP method to the CNLP form will produce the same iterates as SLCP (with unit steps) applied to the NLCP *provided* that global minima are found for indefinite QP subproblems. If a unit step cannot be taken, the equivalence is maintained simply by updating the multiplier estimates in the same manner as the primal iterates.

To avoid potential confusion, it is perhaps important to distinguish this result from a seemingly similar observation of Josephy in [Jo-N 79], which is summarized in Robinson's survey [Rob 82]. His observation concerns applying the generalized equation form of Newton's method to the *special* NLCP that represents the optimality conditions for a *general* NLP. He demonstrates that the method produces a sequence of (bisymmetric) LCPs, each of which represents the optimality conditions for a corresponding member of the sequence of QPs that is generated by applying Wilson's SQP method to the given NLP. This is a result which we will use in the next chapter, but it is altogether different from the equivalence result we obtained above, which pertains to a *general* NLCP and the *special* CNLP constructed to be equivalent to it.

4.4 A sequence of linear programs (SLP)

Perhaps the earliest practical method for solving sizeable nonlinear programs is based on a sequence of linear programs. Origination of the method is attributed to Griffith and Stewart [GS 61], and a recent robust implementation is reported in [ZKL 85]. The LPs are defined by using first-order Taylor's expansions of the objective and constraint functions. Additional bounds are usually added to the subproblems so as to prevent large deviations from the linearization point. These bounds create a so-called *trust region* over which the linearization is presumed to be an acceptably accurate approximation of the nonlinear functions. The bounds further serve to move the subproblem solutions away from basic solutions of the linearized constraints. For the general NLP, we would not expect an optimal solution to be such a basic solution. For a CNLP problem like Eq-NLP, however, we know that basic optimal (complementary) solutions do in fact exist (given the rank condition discussed in Section 4.1.1). In such a case, we can reasonably consider defining LP subproblems without any additional bounds (and worry about convergence later).

For application to problem Eq-NLP, we define the following subproblem at linearization point p^k:

$$\boxed{\text{LP}(p^k)}$$

$$\begin{array}{lll} & \text{minimize} & v \\ & \text{subject to} & \\ \text{(LP1)} & -\nabla d(p^k)p + Ay + hv \geq d(p^k) & \perp \quad \hat{p} \geq 0 \\ \text{(LP2)} & -A^{\mathsf{T}}p \geq 0 & \perp \quad \hat{y} \geq 0 \\ \text{(LP3)} & -h^{\mathsf{T}}p = -1 & \perp \quad \hat{v} \\ \text{(LP4)} & p \;,\; y \geq 0 & \end{array}$$

$\text{LP}(p^k)$ is identical to $\text{QP}(p^k)$ except for the use of a linear objective function. In this sense, SLP may be considered a special case of SQP that uses a vacuous (and therefore positive semidefinite) approximation of the Hessian of the Lagrangian. The iterative process is also the same.

If h is chosen (as described in Chapter 6) to guarantee the feasibility and solvability of any linearized subproblem, then both the primal and dual of $\text{LP}(p^k)$ must be feasible. Consequently, by weak duality, an optimal solution must exist.

If $\bar{p}$ is an equilibrium price vector, subproblem $\text{LP}(\bar{p})$ is the linear program alluded to in Observation 2 of Section 3.3.1 for which the complementary slackness conditions are equivalent to the optimality conditions for problem Eq-NLP.

4.5 A projected (augmented) Lagrangian method

The projected Lagrangian algorithm has proved to be an effective means for solving large, mostly linear optimization problems. It is thus a natural candidate for solving problem Eq-NLP. The method was originated by Robinson [Rob 72] and Rosen and Kreuser [RK 72], and has been implemented in the optimization package MINOS by Murtagh and Saunders [MS 82]. This algorithm also solves a sequence of linearly constrained subproblems, based on Taylor's expansions of the nonlinear functions. (As these constraints are identical to those of the previous three methods, we will not delineate them again here.) The distinctive feature of the projected Lagrangian method is the objective function, which incorporates the original problem objective (as is) and a Lagrangian term based on an estimate of the optimal multipliers and the difference between the linearized and actual constraint function values. As implemented in MINOS, the objective optionally includes an additional quadratic penalty term (also based on the linearization errors) designed to stabilize the iterative process and promote convergence from any starting point. This addition corresponds to the quadratic term of an augmented Lagrangian merit or penalty function.

As applied to Eq-NLP, the objective takes the following form at iteration $k+1$. Let ρ be a penalty parameter, μ^k be an estimate of the optimal Lagrange multipliers for the nonlinear inequalities (NLP1), and recall $d^k(p)$ as a notation for the first-order Taylor's expansion of $d(p)$ at p^k. The objective is then:

$$\text{minimize} \quad v - (\mu^k)^{\mathsf{T}} \left[d^k(p) - d(p) \right] + \tfrac{1}{2}\rho \left\| d^k(p) - d(p) \right\|_2^2 .$$

Given a solution of the subproblem, p^k and μ^k are updated as for SQP with the step length determined by a linesearch procedure.

Typically, the multiplier estimate μ^k is taken to be $\hat{p}^k$, the multipliers for the linearized constraints in the subproblem solution. We have found that two deviations from this procedure yield interesting connections to the solution methods previously described. In both cases we must take $\rho = 0$, which is generally safe for a well-behaved problem. If μ^k is always taken to be zero, the resulting subproblem is precisely SLP(p^k). This is a general result for any NLP with a linear objective function. As an alternative, intuition suggests the possibility of using the prior knowledge that the prices, p, solved for in the subproblem are also good estimates of the optimal multipliers. If we then replace the *constant* μ^k with the *variable* p, the now "endogenous" Lagrangian term transforms the objective function to:

$$v - p^{\mathsf{T}} \left[d^k(p) - d(p) \right] = v - p^{\mathsf{T}} d^k(p) - 0 = v - p^{\mathsf{T}} \nabla d(p^k) p - p^{\mathsf{T}} d(p^k),$$

where again we apply Walras' law, (D2) of Section 2.1. The rightmost expression may be recognized as the non-symmetric form of the objective function for Wilson's SQP(p^k), which is in turn the complementarity conditions for SLCP(p^k).

4.6 Other linearization methods

Another diverse class of solution methods for general nonlinear complementarity problems uses linearizations other than first-order Taylor's expansions. For concreteness, denote by

B^k a matrix depending on the iterate p^k such that:

$$d(p) \approx d(p^k) + B^k(p - p^k).$$

By substituting this linearization into (LCP1) above, an LCP for iteration $k+1$ is obtained, and an iterative procedure otherwise identical to SLCP can be pursued. This class of methods includes the quasi-Newton method for generalized equations as studied by Josephy [Jo-Q 79] and all of the linear approximation methods for complementarity problems surveyed by Pang and Chan [PC 82].

There are two possible reasons in general for considering an alternative linearization scheme. The first is to avoid explicit computation of the Jacobian matrix if the derivatives are expensive to evaluate. In this case, B^k must be substantially easier to calculate than $\nabla d(p^k)$. In a CGE model, particularly one with a sizeable linear production component, it is not in general likely that the evaluation of a Jacobian once each iteration will loom large relative to the computational effort involved in processing the linearized subproblem. There are always exceptions, of course, but we will not consider situations in which expensive derivatives effectively rule out the application of first-order methods.

The other reason for using an alternative linearization is choosing a form of B^k which produces a subproblem that is easier to solve than LCP(p^k). This is a much more relevant issue in the context of problem Eq-NLCP, for which $\nabla d(p^k)$ is singular and has no demonstrably desirable computational properties. While each subproblem can be successfully processed by Lemke's method *regardless* of the linearization employed (see Chapter 6), other linearizations yield subproblems amenable to solution by other, possibly faster, methods. For instance, using $B^k = \nabla d(p^k) - \nabla^T d(p^k)$ has the interesting property that the linearization satisfies Walras' law, (D2) of Section 2.1. The skew-symmetry further implies that each LCP subproblem, expressed as its associated composite quadratic program, is actually a linear program because the Hessian vanishes. Using a general positive semidefinite B^k yields a subproblem whose composite QP can be effectively solved by a convex QP code as well as by any number of iterative (non-pivoting) methods. Finally, using a symmetric positive semidefinite B^k produces a bisymmetric LCP subproblem that comprises the optimality conditions for a convex QP in the price space alone. This lower dimensional QP can in general be solved much faster than LCP(p^k).

As is to be expected, the advantages of easier subproblems have to be weighed against inferior convergence properties of the overall process. These convergence issues are the subject of the next chapter.

5

CONVERGENCE PROPERTIES OF SOLUTION METHODS

In this chapter we present a number of observations about the convergence properties of the methods discussed in the previous chapter for solving problems Eq-NLCP and Eq-NLP. Three related but separate convergence issues are relevant. All involve some notion of a target point. In the case of a complementarity problem, target points are unambiguously the complementary solutions. In an optimization context, target points are generally taken (for purposes of convergence analysis) to be local optimizers, which in the absence of (some generalization of) convexity need not be global optimizers.

The first issue is generally referred to as *local* convergence. The question here is whether there is a neighborhood around a target point such that, if the iterates generated by an algorithm enter that neighborhood, all subsequent iterates remain in the neighborhood and moreover converge to the target point. An important associated issue is the rate at which the iterates converge to the target point once in this neighborhood. The next issue is widely referred to as *global* convergence. The question is whether, from any starting point, the iterates generated by the algorithm eventually converge to some target point. Herein arises an unfortunate ambiguity in the established terminology, since global convergence does not mean convergence to a global optimum. Locating a global optimum is precisely our third convergence issue, which is relevant only in applying optimization methods to Eq-NLP. It is important that converging iterates in fact approach a global minimum, since it is global minima that correspond to complementary (equilibrium) solutions.

5.1 Local convergence

In light of the equivalence between SLCP and Wilson's SQP method in an equilibrium context, we provide below an analysis of the relationship between the known local convergence properties of the two methods. For completeness, we will also briefly survey known results for the other solution methods we have identified in Chapter 4.

5.1.1 SLCP and Wilson's SQP method

Recall the equivalence discussed in Section 4.3 between SLCP and Wilson's SQP method as applied to the nonlinear complementarity problem and its CNLP nonlinear programming counterpart. In light of this equivalence, it seems sensible to invoke the most powerful of the convergence results available for the two methods. Using some precursors of the theory of generalized equations, Robinson in [Rob 74] establishes the local quadratic convergence of a family of optimization methods that includes Wilson's SQP algorithm. His result requires three conditions for a local minimizer to possess a domain of attraction:

1. second-order sufficient optimality conditions

2. linear independence of the gradients of the binding constraints (including simple nonnegativity constraints)
3. strict complementary slackness.

These are strong conditions, but little exists to this day in the way of convergence results for optimization methods without similar conditions.

Josephy in [Jo-N 79] establishes the local quadratic convergence of Newton's method for generalized equations. In particular, any regular solution posseses a domain of attraction. Robinson develops conditions for the regularity of a solution of an NLCP in [Rob 80]. (These conditions were delineated in Section 4.1.1.) He also demonstrates regularity of a solution of an NLP (without strict complementarity) under a strengthening of the second-order sufficiency condition and linear independence of the binding constraint gradients. In later work [Rob 82], he observes that these latter conditions are sufficient for local quadratic convergence of Wilson's method. This follows from Josephy's demonstration of the equivalence between Wilson's method applied to a given NLP and Newton's method applied to the NLCP formed from the optimality conditions for the NLP. (Recall our discussion in Section 4.3 that this is not the same equivalence as the one we develop.)

Unfortunately, this strengthening of the convergence result for Wilson's method on general problems is of no help in the context of applying Wilson's method to the CNLP form of a given NLCP (these forms were defined in Section 3.1). This is because, as somewhat casually observed by Kyparisis in [Kyp 86], the linear independence condition cannot be satisfied in the absence of strict complementarity (or nondegeneracy in the NLCP context). By way of a quick demonstration, we can use the general form of an NLCP since the result is general. Define the usual index sets for a complementary point z:

$$J = \{i : f_i(z) = 0, z_i > 0\},\ K = \{i : f_i(z) = 0, z_i = 0\},\ L = \{i : f_i(z) > 0, z_i = 0\}.$$

Also, let $M \equiv \nabla f(z)$ and let I be a comformably dimensioned identity matrix. The matrix of binding constraints including the nonnegativity conditions is:

$$\left[\begin{array}{c|cc} M_{JJ} & M_{JK} & M_{JL} \\ M_{KJ} & M_{KK} & M_{KL} \\ \hline & I_{KK} & \\ & & I_{LL} \end{array}\right]$$

Clearly, the rows of this matrix are linearly independent if and only if the northwest submatrix has full row rank. But this requires that M_{JJ} be nonsingular and that $K = \emptyset$.

In light of this situation, the equivalence we establish between SLCP and Wilson's SQP method is of no particular help in investigating local convergence. In our complementarity context, the convergence conditions for Wilson's method turn out to be a special case of the regularity condition needed to establish the local quadratic convergence of Newton's method. Recall Robinson's result (discussed in Section 4.1.1) that a solution of an NLCP is regular if and only if M_{JJ} is nonsingular and either $K = \emptyset$ or all principal minors of

the Schur complement, $M_{KK} - M_{KJ}M_{JJ}^{-1}M_{JK}$, are positive. Of course, any regular solution must be locally unique, so any regular solution of an NLCP with $K = \emptyset$ must satisfy the three strong conditions given above for the equivalent CNLP problem.

An interesting by-product of this otherwise disappointing result is that it identifies a class of problems for which Wilson's method will exhibit local quadratic convergence under weaker conditions than linear independence and strict complementarity. Thus, Robinson's sufficient conditions for Wilson's method cannot be necessary conditions as well.

Returning to our central theme, the useable results from these observations are that we can indeed expect local quadratic convergence from SLCP (Wilson's SQP) when the model has regular equilibria *and* the global path of the algorithm brings the iterates into some domain of attraction.

5.1.2 Other optimization methods

Practical implementations of SQP typically combine a positive definite approximation of the Hessian of the Lagrangian with a merit function of some kind to guage descent. The issue of local convergence then becomes inseparable from the mechanism used to promote global convergence. Local superlinear convergence can be achieved if the Hessian approximation approaches the true Hessian in a certain way and the line search permits unit steps in the neighborhood of a local minimum. The precise conditions are unimportant for present purposes, but a summary with references can be found in [GMSW 87].

The SLP method is in essence choosing a steepest-descent search direction (subject to some trust region bounds) with respect to an absolute value (or ℓ_1) penalty function representation of the given NLP. Consequently, we cannot expect better than linear convergence on a general problem. If the problem at hand has a solution at a vertex of the linearized constraints, however, it is an intuitive result that, once the correct basis has been determined, the method is actually performing Newton's method on the square system of basic variables and binding constraints. We could then expect quadratic convergence in the neighborhood of a regular solution. Zhang, Kim and Lasdon assert this result (without proof) in [ZKL 85].

The projected Lagrangian method was shown by Robinson [Rob 74] to be locally quadratically convergent under the same conditions as given above for Wilson's SQP method. The practical implementation in MINOS inherits the same property, since the penalty parameter ρ is decreased to zero as a local minimum is approached.

5.1.3 Other linearization methods

Pang and Chan in [PC 82] unify an extensive collection of convergence results for linear approximation methods for solving variational inequalities, organized around the basic themes of norm-contraction, vector-contraction, and monotonicity. Further elaboration on these results is not justified in our context, since the general equilibrium problem does not demonstrably satisfy any of the known conditions for local (or global) convergence.

Josephy in [Jo-Q 79] extends the known results for solving nonlinear equations by quasi-Newton methods to the case of generalized equations. In particular he shows that, under conditions satisfied by the usual quasi-Newton update formulas, the sequence of iterates is linearly convergent in the neighborhood of a regular solution. Of incidental interest is the condition for superlinear convergence of a (known-to-be) linearly convergent sequence of quasi-Newton iterates. In our context, this specializes to:

$$\lim_{k\to\infty} \frac{\left\|[B^k - \nabla d(p^k)](p^k - p^{k-1})\right\|}{\|p^k - p^{k-1}\|} = 0,$$

where B^k is the approximation matrix at iteration k.

5.2 Global convergence

We first briefly survey the known global convergence properties of the solution methods we have identified. We then develop some new machinery for analyzing convergence in the context of an equilibrium or complementarity problem.

In the absence of one of the contraction or monotonicity properties discussed by Pang and Chan, very little is known about the global convergence properties of Newton, quasi-Newton, or other linearization methods for solving the NLCP. The same can be said about Wilson's SQP method in the absence of convexity. As an empirical matter, however, Mathiesen's computational experience has shown SLCP to be remarkably robust.

Global convergence has been proved for well-designed SQP methods, given the usual assumptions about a strict local minimizer with strict complementarity and linear independence of the gradients of the binding constraints. (Recall that linear independence and strict complementarity are inseparable for a problem of type CNLP.) "Globalization" has been based on line-search procedures relative to an absolute value (ℓ_1) merit function or to an augmented Lagrangian merit function. In either case, the convergence argument depends critically on the positive definiteness of the matrix used to represent the Hessian of the Lagrangian. Again see [GMSW 87] for a summary of these results and references to the detailed demonstrations.

Zhang, Kim and Lasdon in [ZKL 85] demonstrate the global convergence of their SLP method by use of an ℓ_1 exact penalty function combined with trust region bounds. Since SLP is a special case of SQP with vacuous Hessian, it is intuitive to think of the tightening of trust region bounds as a substitute (in the convergence argument) for positive definiteness of the estimated Hessian. How best to determine satisfactory penalty weights is still unresolved.

Despite the close connection between the objective function used in MINOS for the linearized subproblems and the augmented Lagrangian merit function used in some SQP procedures, there has yet to be a rigorous theoretical demonstration of the global convergence of the algorithm. The method has strong intuitive appeal, however, and a wide range of successful applications to nonlinear problems has been reported, both in the literature and informally.

5.2.1 A differentiable exact penalty function (Eq-DEPF)

In the course of examining the ℓ_1 and augmented Lagrangian penalty functions (particularly with respect to their use as merit functions for global convergence analysis), we discovered a potentially fruitful specialization of the approach to problem Eq-NLP. This specialization provides, perhaps for the first time, an unambiguous means of evaluating the progress of a sequential method towards an equilibrium solution. While both penalty functions specialize in an identical fashion, we prefer to deal with the differentiable augmented Lagrangian function.

The specialization begins by replacing the $d(p)$ term in inequality (NLP1) of Section 3.3 with a proxy variable x (which is not constrained in sign). Instead of adding a constraint, $x = d(p)$, we add to the previously linear objective function a Lagrangian term and a quadratic penalty term, i.e.,

$$-\mu[x - d(p)] + \tfrac{1}{2}\rho \|x - d(p)\|_2^2 .$$

In the typical case, μ would be an iteratively updated estimate of the Lagrange multipliers for the omitted constraint, $x = d(p)$. Instead, as we illustrated in Section 4.5, we can use our prior knowledge of the desired complementarity relationships and replace the updated constant μ with the variable p. The result is a *linearly constrained* problem which we will show to be equivalent to Eq-NLP (and thus to Eq-NLCP as well).

Eq-DEPF

$$\begin{array}{lrcccccccl}
 & \text{minimize} & -p^{\mathsf{T}}x + \tfrac{1}{2}\rho\|x-d(p)\|_2^2 & + & v & & & & & \\
 & \text{subject to} & & & & & & & & \\
\text{(DEPF1)} & & -x & + & Ay & + & hv & \geq & 0 & \perp \quad \hat{p} \geq 0 \\
\text{(DEPF2)} & & -A^{\mathsf{T}}p & & & & & \geq & 0 & \perp \quad \hat{y} \geq 0 \\
\text{(DEPF3)} & & -h^{\mathsf{T}}p & & & & & = & -1 & \perp \quad \hat{v} \\
\text{(DEPF4)} & & p & , & y & & & \geq & 0 &
\end{array}$$

Observe first that $v - p^{\mathsf{T}}x \geq 0$ for any feasible solution. By the usual device, multiply through inequalities (DEPF1) by $p \geq 0$ and (DEPF2) by $y \geq 0$. Add the resulting two inequalities and $v - p^{\mathsf{T}}x \geq 0$ follows. Since the penalty term is nonnegative, we can conclude that the objective function is bounded below by zero on the feasible region.

Equivalence between Eq-DEPF and Eq-NLP is then easily established. Given any global minimum solution of Eq-NLP (i.e., $(\bar{p}, \bar{y}, \bar{v})$ feasible with $\bar{v} = 0$), we immediately construct a global minimum of Eq-DEPF as $(\bar{p}, \bar{x}, \bar{y}, \bar{v})$ with $\bar{x} = d(\bar{p})$. Note that the penalty term vanishes and $\bar{p}^{\mathsf{T}}\bar{x} = \bar{p}^{\mathsf{T}}d(\bar{p}) = 0$ by virtue of Walras' law, (D2) of Section 2.1. We thus know that the globally minimal objective value for Eq-DEPF is in fact zero. Given this, we can argue the converse. Any global minimum $(\bar{p}, \bar{x}, \bar{y}, \bar{v})$ for Eq-DEPF must have $\bar{x} = d(\bar{p})$ or

else the penalty term would be positive. In this case, we again have $\bar{p}^T\bar{x} = \bar{p}^T d(\bar{p}) = 0$ by virtue of Walras' law. It then follows that $\bar{v} = 0$ since the global minimum objective value is zero. We then immediately have $(\bar{p}, \bar{y}, \bar{v})$ feasible with $\bar{v} = 0$ as a global minimum solution for Eq-NLP.

Note that the above equivalence holds for *any* value of $\rho > 0$. This is because $v - p^T x \geq 0$ on the feasible region of Eq-DEPF and zero is in fact an attainable global minimum. This invariance to ρ is in happy contrast to the usual situation with penalty functions, in which the correspondence between optimal solutions of the penalty problem and optimal solutions of the original constrained problem obtains only for sufficiently large values of the penalty parameter. We are thus free in problem Eq-DEPF to vary the penalty parameter in any way conducive to an analysis of global convergence.

Any number of further transformations of problem Eq-DEPF are possible that still maintain the equivalence to problem Eq-NLP. Other types of penalty terms could be used, differentiable or otherwise. Since feasibility of subproblems is not an issue, variable v could be dispensed with altogether. Having omitted v, the term $-p^T x$ could be dropped from the objective function, leaving only the penalty term. (This clearly highlights the underlying character of the equilibrium problem as a feasibility problem.) The disadvantage of this transformation from a computational perspective is that the resulting variant of Eq-DEPF would have zero Lagrange multipliers at the global minimum. In contrast, a global minimum of Eq-DEPF as stated has the desirable property that the "primal" variables are also complementary Lagrange multipliers. Leaving v in Eq-DEPF also allows a more direct and convenient correspondence to the solutions of linearized subproblems in which the presence of v is used to guarantee the existence of feasible and complementary solutions. (We discuss this further in Chapter 6.)

The significance of problem Eq-DEPF is twofold. First, the value of the objective function can be used as a merit function for judging the progress of any sequential method that produces a (p, x, y, v) which is feasible for Eq-DEPF. All of the solution methods that we discuss indeed have this property; x is directly obtained as the value of the linearized demand function in the solution of the subproblem. Second, it may be viable to apply an optimization method directly to problem Eq-DEPF instead of problem Eq-NLP. Both of these uses of Eq-DEPF will be discussed in turn below, but first a brief aside on the immediate generalization of this penalty function formulation to problem CNLP.

A more general result

Similar to what we discovered in Section 4.3, a more general result lurks behind the developments for problems Eq-NLCP and Eq-NLP. A differentiable exact penalty function representation is also readily available for the equivalent CNLP form of the general nonlinear complementarity problem (see Section 3.1). It has the following simple form:

$$\text{(DEPF)} \quad \text{minimize} \quad w^T z + \tfrac{1}{2}\rho \|w - f(z)\|_2^2 \quad \text{subject to} \quad w \geq 0,\ z \geq 0.$$

Again, we have a linearly constrained problem with an objective function that is bounded below by zero on the feasible region (the nonnegative orthant in this case). Assuming

that the original NLCP has a solution, any such solution is a global minimum of DEPF and conversely. As in the case of Eq-DEPF, any linear parts of $f(z)$ can optionally be maintained as general linear constraints as opposed to being included in the penalty term.

Analogous to the situation with the specific Eq-DEPF problem, formulation DEPF may prove to have independent computational potential as well as usefulness in evaluating the convergence of any sequential method for solving an NLCP.

5.2.2 Stationary points of Eq-DEPF

Problem Eq-DEPF inherits the nonconvexity of problem Eq-NLP, but it also inherits the interesting properties of the first-order optimality conditions which suggest that nonoptimal stationary points may be rare or avoidable. These conditions are thus worth delineating here. Let $(\bar{p}, \bar{x}, \bar{y}, \bar{v})$ be a Karush-Kuhn-Tucker point of Eq-DEPF with associated Lagrange multipliers $(\hat{p}, \hat{y}, \hat{v})$. The primal conditions are simply (DEPF1)–(DEPF4) evaluated at $(\bar{p}, \bar{x}, \bar{y}, \bar{v})$. The multiplier conditions are as follows:

$$
\begin{array}{lrcccccccccl}
\text{(DMx)} & \rho[\bar{x}-d(\bar{p})] & - & \bar{p} & + & \hat{p} & & & = & 0 & \perp & \bar{x} \\
\text{(DM0)} & -\rho\nabla^{\mathsf{T}}d(\bar{p})\,[\bar{x}-d(\bar{p})] & - & \bar{x} & & & + A\hat{y} + h\hat{v} & \geq & 0 & \perp & \bar{p} \geq 0 \\
\text{(DM1)} & -[\bar{x}-d(\bar{p})] & + & \nabla^{\mathsf{T}}d(\bar{p})\hat{p} & + & A\hat{y} & + h\hat{v} & \geq & 0 & \perp & \bar{p} \geq 0 \\
\text{(DM2)} & & & -A^{\mathsf{T}}\hat{p} & & & & \geq & 0 & \perp & \bar{y} \geq 0 \\
\text{(DM3)} & & & -h^{\mathsf{T}}\hat{p} & & & & = & -1 & \perp & \bar{v} \\
\text{(DM4)} & & & \hat{p} & , & \hat{y} & & \geq & 0 & &
\end{array}
$$

Here, we have obtained (DM1) from the "original" condition (DM0) by applying the identity (DMx). Conditions (DM1)–(DM4) then differ from the Eq-NLP multiplier conditions (M1)–(M4) (see Section 3.3.1) only in the presence of the term $[\bar{x}-d(\bar{p})]$. Equation (DMx) also adds a direct connection between $\bar{p}$ and $\hat{p}$ that is not present in the KKT conditions for problem Eq-NLP.

We can now detail a number of interesting observations about these conditions.

Observation 1. By the usual manipulations of the complementarity conditions, we derive that $\bar{v} = \hat{p}^{\mathsf{T}}\bar{x}$ and $\hat{v} = \bar{p}^{\mathsf{T}}\bar{x}$.

Observation 2. If $\bar{x} = d(\bar{p})$, then $\hat{p} = \bar{p}$ by (DMx). In turn we must have $\bar{v} = \hat{v} = 0$ by Walras' law, property (D2) of Section 2.1. In short, we have an equilibrium solution.

Observation 3. Since $v - p^{\mathsf{T}}x \geq 0$ for any feasible solution of Eq-DEPF, we can use Observation 1 and (DMx) to conclude:

$$0 \leq \bar{v} - \bar{p}^{\mathsf{T}}\bar{x} = \bar{v} - \hat{v} = (\hat{p} - \bar{p})^{\mathsf{T}}\bar{x} = -\rho[\bar{x} - d(\bar{p})]^{\mathsf{T}}\bar{x}.$$

If $\bar{x} \neq d(\bar{p})$, then $\|\bar{x} - d(\bar{p})\|_2^2 > 0$ requires:

$$\rho[\bar{x} - d(\bar{p})]^{\mathsf{T}} d(\bar{p}) < 0 \implies (\bar{p} - \hat{p})^{\mathsf{T}} d(\bar{p}) < 0 \implies \hat{p}^{\mathsf{T}} d(\bar{p}) > 0,$$

where the first implication follows from (DMx) and the second from Walras' law. In Observation 4 of Section 3.3.1, we obtained the same conclusion for a nonoptimal stationary point of Eq-NLP.

Observation 4. Given the price normalizations (DEPF3) and (DM3), equation (DMx) implies that $h^{\mathsf{T}}[\bar{x} - d(\bar{p})] = 0$. Since the normalization bounds individual prices as well, it also follows that $\rho h_i |\bar{x}_i - d_i(\bar{p})| \leq 1$. Thus, if $h > 0$ the deviations $\bar{x} - d(\bar{p})$ obtained at a KKT point of Eq-DEPF can be made arbitrarily small by increasing ρ. Intuitively, it is hard to believe that a non-contrived problem could have such near-equilibrium stationary points sufficiently isolated from true equilibria that a sequence of iterates would be attracted to a nonoptimal point rather than to an equilibrium point.

We will have further use for some of these conditions and observations at the end of this chapter.

5.2.3 Judging descent using Eq-DEPF

For brevity, let $\Phi(p, x, v)$ denote the objective function of Eq-DEPF. We will need its gradient:

$$\nabla\Phi(p^k, x^k, v^k) = \left[-x^k - \rho\nabla^{\mathsf{T}} d(p^k)\left[x^k - d(p^k)\right],\ -p^k + \rho\left[x^k - d(p^k)\right],\ 1\right]^{\mathsf{T}}.$$

Recall that the general sequential procedure for all the methods we discuss begins at iteration $k+1$ by linearizing the demand functions $d(p)$ at the point p^k. It will help to have two shorthand notations for the possible linearizations:

$$\begin{aligned} d^k(p) &\equiv d(p^k) + \nabla d(p^k)(p - p^k) = d(p^k) + \nabla d(p^k)p \\ b^k(p) &\equiv d(p^k) + B^k(p - p^k). \end{aligned}$$

Note that $b^k(p)$ includes the Taylor's expansion $d^k(p)$ as a special case. We will use the non-specific $b^k(p)$ when an expression or remark applies equally well to any linearization, and will explicitly call notice to any reference that expressly excludes the Taylor's expansion.

We will also use SP(p^k) to denote the subproblem defined at iteration $k+1$ when it is not important to distinguish between specific methods.

If $(\bar{p}, \bar{y}, \bar{v})$ solves SP(p^k), we take $\bar{x} = b^k(\bar{p})$ and immediately obtain a feasible solution of Eq-DEPF. Its objective value can be readily compared to those of previous iterations. If the solution obtained for SP(p^k) is a complementary solution, then $\bar{v} - \bar{p}^{\mathsf{T}} b^k(\bar{p}) = 0$, since this expression is precisely the complementarity conditions for the subproblem. In this case, $\Phi(\bar{p}, \bar{x}, \bar{v}) = \frac{1}{2}\rho\|\bar{x} - d(\bar{p})\|_2^2$. The interesting implication here is that, in comparing two successive iterates of the SLCP procedure (for instance), all that really matters is

the accuracy of the linear approximation to the demand functions. Counter to prevailing intuition, the combinatorial aspects of identifying basic production activities and binding inequality constraints are important only insofar as they affect the prices at which the linearized demand functions are evaluated.

We can also use the subproblem solution to determine a search direction, i.e., $(\bar{p}-p^k, \bar{x}-x^k, \bar{y}-y^k, \bar{v}-v^k)$. This direction is always locally feasible and will furthermore be a descent direction relative to Eq-DEPF if:

$$\nabla\Phi(p^k, x^k, v^k)(\bar{p}-p^k, \bar{x}-x^k, \bar{v}-v^k) < 0.$$

Then, in principle at least, a line search could be performed with respect to $\Phi(p, x, v)$ and an appropriate step length chosen ($0 \leq \lambda \leq 1$). Note that $x^{k+1} = b^k(p^{k+1})$ only if $\lambda = 1$.

We can now examine the search direction for descent.

$$\begin{aligned}
\Phi' \equiv \nabla\Phi(p^k, x^k, v^k)(\bar{p}-p^k, \bar{x}-x^k, \bar{v}-v^k) \quad &= \quad \text{(upon some reordering)} \\
(\bar{v}-v^k) - (p^k)^{\mathsf{T}}(\bar{x}-x^k) - (x^k)^{\mathsf{T}}(\bar{p}-p^k) - \rho\left[x^k - d(p^k)\right]^{\mathsf{T}}\left\{-\bar{x} + x^k + \nabla d(p^k)(\bar{p}-p^k)\right\} & \\
= \left(\bar{v} - (p^k)^{\mathsf{T}}\bar{x}\right) + \left(v^k - \bar{p}^{\mathsf{T}}x^k\right) - 2\left(v^k - (p^k)^{\mathsf{T}}x^k\right) & \\
- \rho\left[x^k - d(p^k)\right]^{\mathsf{T}}\left\{\left[x^k - d(p^k)\right] - \left[\bar{x} - d^k(\bar{p})\right]\right\} &
\end{aligned}$$

Here, the transformation of the first three terms is by reordering and by adding and subtracting v^k. The transformation of the sub-term in braces is by adding and subtracting $d(p^k)$ and then substituting in the shorthand $d^k(\bar{p})$. Note that we will use Φ' as an abbreviation for this descent formula.

Since both $(p^k, \bar{x}, \bar{y}, \bar{v})$ and $(\bar{p}, x^k, y^k, v^k)$ are feasible solutions for Eq-DEPF, it follows that the first two terms of Φ' are nonnegative. Similarly, the third term is nonpositive. *If* $x^k = b^{k-1}(p^k)$ (that is a unit step was taken at the previous iteration) *and* a complementary solution was obtained to subproblem $\mathrm{SP}(p^{k-1})$, then we can conclude that the third term vanishes because (as mentioned above) it is the statement of the complementarity conditions for $\mathrm{SP}(p^{k-1})$. Other than this, we are unable to state anything definitive about the magnitudes of these first three terms.

The fourth term indicates the leverage that can be exercised by the choice of the penalty parameter ρ. For the four classes of methods that linearize by means of the Taylor's expansion, $\bar{x} = d^k(\bar{p})$. In this case the fourth term reduces to $-\rho\left\|x^k - d(p^k)\right\|_2^2$. Clearly, this quantity is strictly negative so long as the iterations have not converged. So the search direction obtained by $\mathrm{SP}(p^k)$ can be considered a descent direction with respect to any manifestation of Eq-DEPF that has a sufficiently large value of ρ. This result depends only on the use of Taylor's expansions; the objective function or complementarity condition for the subproblem does not matter. The nature of the solution obtained by the subproblem affects only the magnitudes of the first three terms, the sum of which conditions how large ρ must be in order to make Φ' negative.

Unfortunately, this somewhat encouraging result does not necessarily apply for linearizations other than the Taylor's expansion. In this case we have:

$$\bar{x} - d^k(\bar{p}) = b^k(\bar{p}) - d^k(\bar{p}) = \left[B^k - \nabla d(p^k)\right](\bar{p} - p^k).$$

(The expression on the right may be recognized as the vector in the numerator of the limit condition for establishing superlinear convergence of quasi-Newton iterates.) The influence of the deviation $\bar{x} - d^k(\bar{p})$ on the sign of the fourth term of Φ' is ambiguous. Should the fourth term be positive, it may or may not be possible to make ρ *small* enough to obtain a local descent. In particular, this will not be possible if the third term vanishes, as indeed it will if a unit step was taken on the previous iteration. (Recall that the non-Taylor linearization methods we discuss all obtain complementary solutions of the subproblems.)

These last two results establish an important connection between local and global convergence conditions. They also deliver a rather serious second blow to the prospects of using other linearizations on a general equilibrium problem. The use of Taylor's expansions (and the corresponding relationship to Newton's method) is central to the *local* convergence arguments for SQP methods, the projected Lagrangian method, and SLP (given a vertex solution). Their use allows also for the superlinear and quadratic local convergence rates of these methods. We now see also that use of Taylor's expansions means that any search direction obtained can be interpreted as a descent direction in Eq-DEPF for large enough values of ρ. So their use is also conducive to *global* convergence of the iterates. In sharp contrast, the use of other linearizations typically dooms local convergence to a linear rate *given* that the iterates converge at all. At the same time departures from Taylor's expansions endanger the assumption of convergence by producing search directions that could in fact be uphill. It is further interesting in this respect that the condition for superlinear local convergence of quasi-Newton iterates is mutually reinforcing with the condition promoting global convergence in the sense that "small" values of $\left[B^k - \nabla d(p^k)\right](\bar{p} - p^k)$ are more likely to produce a negative fourth term in Φ'.

We have been careful not to assert that any of the above proves the global convergence of methods using Taylor's expansions. It merely provides a partial theoretical explanation for the empirical record, particularly as regards SLCP. Two issues continue to elude resolution: the boundedness of ρ and the choice of step length.

Boundedness of the values of ρ required to indicate descent is not strictly necessary for global convergence. These values can grow indefinitely provided a number of other stringent conditions are met, including for example:

$$\Phi' \leq -\gamma \left\| \nabla\Phi(p^k, x^k, v^k) \right\| \left\| (\bar{p} - p^k, \bar{x} - x^k, \bar{v} - v^k) \right\|,$$

where $\gamma > 0$. In the literature on SQP and SLP, verification that the iterates satisfy such conditions is critically dependent on either the use of positive definite estimated Hessians or sufficiently small trust regions. We have not been able to demonstrate that the special properties of an equilibrium problem in some way obviate the necessity of such mechanisms.

Given a downhill search direction, it also remains unclear whether a step length chosen by any means simpler than a line search would necessarily yield a net descent. We are

particularly concerned here about the unit steps taken whenever possible by SLCP (or the other linearization methods) and the deviations from this if the unit step leads outside the domain of $d(p)$.

One final, perhaps obvious, remark should be made. To the extent that the merit functions routinely employed in optimization algorithms are good approximations of the exact merit function available from Eq-DEPF, it is reasonable to expect that descent directions and step lengths determined by those algorithms will also be valid with respect to Eq-DEPF.

5.2.4 Solving Eq-DEPF directly

The relative merits of solving Eq-DEPF directly (as opposed to Eq-NLP) depend both on the size and structure of the specific problem and the optimization method to be employed. A decided advantage of solving Eq-DEPF is avoiding the solution of a difficult subproblem which is only approximate. It also avoids the necessity (in solving Eq-NLP) of having to update or recalculate the factorization of the linearized constraints at each new linearization point. Search directions calculated with respect to the Eq-DEPF objective function have more accurate information about curvature, and the demand estimate x is allowed to move more flexibly than in the linear path $b^k(p)$ available to subproblem SP(p^k) of Eq-NLP.

To be weighed against these advantages is the increase in problem size, both in terms of the matrix of linear constraints and (in a second-order method) the approximation of the projected Hessian that must be maintained (to be denoted by W). In current SQP implementations, the constraint matrix is stored in dense form, while in MINOS the constraints are stored in sparse form. W is in general dense and is therefore maintained in dense form in both SQP routines and MINOS.

In both areas the effects on problem size of adding x are critically dependent on the number of final commodities relative to the number of primary and intermediate commodities. Since demand is constant for primary and intermediate commodities, the vector x need only correspond to the final commodities.

Regardless of the dimension of x, adding x to the problem would not significantly increase the effort of maintaining and factorizing the constraints in MINOS. Indeed, there would be fewer nonzero coefficients than in SP(p^k), since presumably the identity matrix associated with x is much sparser than $\nabla d(p^k)$. Moreover, since the identity columns associated with x must be basic (because x is unconstrained in sign), the basis will be sparser since these columns will displace some columns of A. In contrast, adding columns to the dense representation of constraints in an SQP routine might be a significant burden, growing with the dimensionality of x.

The most serious penalty, however, is with respect to the dimensionality of the (approximate) projected Hessian, which equals the dimension of the null space of the binding constraints. (Problem Eq-DEPF does not inherit the vertex solution property of Eq-NLP.) Each component of x adds a dimension to this null space unless the associated commodity balance happens to be slack — which is a rare occurrence for final commodities. In the context of MINOS, this has the nice interpretation that the dimensionality of W is increased

by each production activity that is displaced from the basis by a component of x. Since W is dense and its factorization must be updated repeatedly, it is unlikely that solving Eq-DEPF directly by a second-order method will be viable for a problem with "many" final commodities. The definition of "many" naturally depends on hardware, computer budget, and user patience.

Maintenance of a projected Hessian approximation can of course be avoided by use of a first-order optimization method, such as steepest feasible descent implemented via SLP (with the trust region bounds in this case). The cost of doing so is an overall linear convergence rate. It is significant to note that, once the objective function of Eq-DEPF is linearized, the problem completely separates into a price problem over constraints (DEPF2)–(DEPF3) and a supply/demand balance problem over (DEPF1). The price problem can then be dualized, leaving two very similar problems on the supply/demand balances.

5.3 Convergence to what?

The methods we have surveyed may be classified into two groups with respect to the kind of solutions they determine. SLCP, the equivalent Wilson's SQP method (given global solutions to the indefinite subproblems), and the other linearization methods compute complementary solutions. In an optimization framework, this amounts to seeking a Karush-Kuhn-Tucker point with the additional qualitative constraint that the Lagrange multipliers equal the "primal" variables. This screening process ensures that if the iterates converge, the point located is a complementary (equilibrium) solution.

In contrast, the other SQP variants, SLP, and the projected Lagrangian algorithms are general optimization procedures that can be applied in principle to either Eq-NLP or Eq-DEPF. They do not presume any special structure, nor can they directly utilize the prior knowledge that the globally optimal objective value is zero. Iterations terminate upon finding any KKT point of the problem. Given the lack of convexity, however, there is no demonstrable guarantee that the stopping point will be a global minimum, i.e., an equilibrium solution.

We discover then the Scylla and Charybdis of solving general equilibrium problems. On one side we have a group of methods which insist on complementary solutions of the linearized subproblems. *If* the iterates converge for one of these methods, they do in fact converge to an equilibrium solution. However, we still have very little in the way of a rigorous reason why they should converge globally from any starting point. On the other side, we have a battery of optimization methods with strong theoretical and/or practical properties which guarantee (or at least make likely) the global convergence of iterates from any starting point. Unfortunately, they may converge to a local minimum rather than to an equilibrium solution. We can readily identify this unsatisfactory termination, but what to do in that event is not at all clear. We make some suggestions relative to this matter in the next subsection.

5.3.1 Descent from nonoptimal stationary points

We have established that global optima of problem Eq-DEPF are in one-to-one correspondence with those of Eq-NLP. There is no particular reason to expect that a similar correspondence would apply to nonoptimal stationary points. This naturally induces the question: if an algorithm applied to one problem terminates at a nonoptimal Karush-Kuhn-Tucker point, is there any perspective available from the other formulation that would suggest a direction of escape? Since any stationary point of Eq-NLP provides a feasible solution of Eq-DEPF, we are in a good position to examine this point from the perspective of Eq-DEPF. We do so below, and exhibit two sensible possibilities for escape—one that yields a local descent and one that does not. In contrast, a stationary point of Eq-DEPF need not provide a feasible solution of Eq-NLP, and meaningful additional perspective can be obtained only from a feasible point. In the case of a jointly feasible point, we can obtain some partial results which we present first.

Stationary points of Eq-DEPF

Suppose that $(\bar{p}, \bar{x}, \bar{y}, \bar{v})$ is a KKT point of Eq-DEPF with associated Lagrange multipliers $(\hat{p}, \hat{y}, \hat{v})$. In light of equation (DMx) from Section 5.2.2, we can conclude that $\hat{p}_i = 0 \implies \bar{x}_i - d_i(\bar{p}) \geq 0$. If $\hat{p}_i > 0$, then the associated inequality of (DEPF1) in Section 5.2.1 must be binding. Consequently, to obtain a feasible solution of Eq-NLP (Section 3.3), we must again have $\bar{x}_i \geq d_i(\bar{p})$. Thus, the necessary and (obviously) sufficient condition for deriving a feasible solution of Eq-NLP from a stationary point of Eq-DEPF is that $\bar{x} \geq d(\bar{p})$.

Recall from Observation 4 of Section 5.2.2 that $h^{\mathsf{T}}[\bar{x} - d(\bar{p})] = 0$. This general notation obscures the fact that in practice we would only define variables x_i for final commodities, i.e., those for which $d_i(p)$ is not a known constant. If $h_i > 0$ for all final commodities i, then the above equality and nonnegativity of $\bar{x}_i - d_i(\bar{p})$ clearly requires $\bar{x}_i = d_i(\bar{p})$. Hence, by virtue of Observation 2 of Section 5.2.2, the price normalization can ensure that any stationary point of Eq-DEPF which is feasible for Eq-NLP must be an optimal (equilibrium) solution. This particular price normalization is precisely the one shown in Chapter 6 to guarantee the feasibility and solvability of all linearized subproblems of Eq-NLP or Eq-NLCP.

Stationary points of Eq-NLP

Suppose that (p^k, y^k, v^k) is a KKT point of Eq-NLP with associated Lagrange multipliers $(\hat{p}, \hat{y}, \hat{v})$. (We use the superscript notation because we will be using the descent formulas of Section 5.2.3.) If the point is nonoptimal, it must be the case that $v^k > 0$ and $\hat{p} \neq p^k$ (recall the discussion of Section 3.3.1). This means that (p^k, y^k, v^k) solves LP(p^k) but not SLCP(p^k), as defined in Sections 4.4 and 4.1, respectively. A feasible solution of Eq-DEPF is immediately obtained by setting $x^k = d(p^k)$. We again investigate search directions of the form $(\bar{p} - p^k, \bar{x} - x^k, \bar{y} - y^k, \bar{v} - v^k)$, where $(\bar{p}, \bar{x}, \bar{y}, \bar{v})$ must be a feasible solution of Eq-DEPF. In this special context, the descent formula simplifies to:

$$\Phi' = \left(\bar{v} - (p^k)^{\mathsf{T}}\bar{x}\right) + \left(v^k - \bar{p}^{\mathsf{T}}x^k\right) - 2v^k.$$

One might intuitively suspect that solving SLCP(p^k) would provide a descent direction when viewed from the perspective of problem Eq-DEPF. Unfortunately, this is not the case. Let $(\bar{p}, \bar{y}, \bar{v})$ solve SLCP(p^k) and set $\bar{x} = d^k(\bar{p})$. Then, using Walras' law (D2) and property (W1) of Section 2.1:

$$(p^k)^T \bar{x} = (p^k)^T \left[d(p^k) + \nabla d(p^k) \bar{p} \right] = -\bar{p}^T d(p^k).$$

Since $x^k = d(p^k)$, we obtain $\Phi' = \bar{v} - v^k$. This looks promising at first, but since (p^k, y^k, v^k) solves LP(p^k) and $(\bar{p}, \bar{y}, \bar{v})$ is feasible for LP(p^k), it must be the case that $\bar{v} \geq v^k$. Hence, the local information indicates that relinearizing at some step in the direction $\bar{p} - p^k$ would not produce a descent in problem Eq-DEPF. This is discouraging but not necessarily terminal. We certainly are not satisfied with the current KKT point, and even if a movement toward $\bar{p}$ does not yield a local descent, it *might* at least initiate a new set of iterates converging to a different KKT point. In particular, following a path of complementary solutions thereafter would ensure that the nonoptimal point would not be revisited.

Another potential search direction is available without further calculation: the Lagrange multipliers from the nonoptimal stationary point of Eq-NLP. If we define $\hat{x} = -\nabla^T d(p^k) \hat{p}$ and examine the multiplier conditions (M1)–(M4) from Section 3.3.1 (with $\bar{p}$ replaced by p^k, etc.), we observe that $(\hat{p}, \hat{x}, \hat{y}, \hat{v})$ is indeed a feasible solution of Eq-DEPF. Now recall from Observations 3 and 4 of Section 3.3.1 that $\hat{v} = 0$ and $v^k = \hat{p}^T d(p^k)$. In the present context, the latter means $v^k = \hat{p}^T x^k$. Also, $(p^k)^T \hat{x} = 0$ because of property (H1) of Section 2.1. The descent formula then reduces to $\Phi' = -2v^k$, which is strictly negative if the stationary point is nonoptimal. In fact, this shows the direction to be a steepest descent direction since it obtains zero for the two nonnegative terms in the formula for Φ'.

Observe that the estimated descent of $-2v^k$ could not be obtained by a unit step in the direction of $(\hat{p}, \hat{x}, \hat{y}, \hat{v})$. This is because $x^k = d(p^k)$ at the current point, which means that the value of the objective is v^k in both Eq-NLP and Eq-DEPF. Since the objective is bounded below by zero, the maximum attainable descent is $-v^k$. A unit step will produce descent if:

$$\begin{aligned} v^k &> -\hat{p}^T \hat{x} + \tfrac{1}{2}\rho \left\| \hat{x} - d(\hat{p}) \right\|_2^2 \\ &= \hat{p}^T \nabla d(p^k) \hat{p} + \tfrac{1}{2}\rho \left\| [\nabla d(\hat{p}) - \nabla d(p^k)]^T \hat{p} \right\|_2^2, \end{aligned}$$

where we have inserted the definition of $\hat{x}$ and applied property (W1) of Section 2.1. We demonstrated in Observation 5 of Section 3.3.1 that the first term of this expression is nonnegative. Hence, it is impossible to say anything a priori as to whether a unit step will necessarily yield a descent.

What we have established, then, is that the Lagrange multipliers for a nonoptimal KKT point of problem Eq-NLP can be used to define a steepest local descent direction from that point when viewed in the context of problem Eq-DEPF. Some step in the direction $\hat{p} - p^k$ may then be used to define a new linearization point. Iterates from that sensible point may then lead to a different KKT point, *provided* once again that something is done to prevent a return to the nonoptimal point.

6

SOLVABILITY OF LINEARIZED SUBPROBLEMS

In this chapter we verify earlier assertions that the normalization vector h can be chosen in such a way as to guarantee the feasibility and solvability of all linearized subproblems encountered in the solution of Eq-NLCP or Eq-NLP. By solvability we mean that all subproblem constraint matrices belong to a class that can be successfully processed by Lemke's almost-complementary pivoting method. In particular, for such matrices Lemke's method can terminate on a secondary ray only if the problem is infeasible. However, the requisite choice of h also guarantees feasibility of the linearized subproblem as long as the original equilibrium problem is feasible. In light of the properties of computable general equilibrium models outlined in Chapter 2, we may safely presume feasibility (and existence of an equilibrium solution) for any properly formulated model. Consequently, the combination of solvability and feasibility guarantees that each linearized subproblem has a complementary solution and that this solution can be computed by at least one method, i.e., Lemke's.

The demonstration of solvability is a direct application of some results of Garcia [Gar 73] on classes of matrices amenable to solution by Lemke's method. His results extend somewhat the earlier results of Eaves [Eav 71] in a way that is remarkably well suited to our purposes. Feasibility will follow immediately from problem structure and the same choice of h as required for solvability.

In previous work, the author [Sto 85] and Eaves [Eav 87] have recognized that choosing $h = e$ and incorporating a matching artificial column guarantees the solvability of the LCP subproblems. Mathiesen in [Mat 87] applies Eaves' earlier results [Eav 71] to a specific small problem in order to investigate the effects of the linearization point and the choice of numéraire on the solvability and existence of solutions to the LCP subproblems. The results obtained below are both more general and justify stronger conclusions.

6.1 Relevant results for Lemke's method

Here we briefly summarize the definitions and results of Garcia [Gar 73] which we will apply in the next section. We have made some minor modifications to notation and wording so as to better fit our context. Use of the vector d and of index sets J and K should not be confused with previous uses in this document.

Denote the standard linear complementarity problem by q/M and the augmented form with an artificial covering vector d by $d/q/M$. That is, we seek a solution of the system:

$$z \geq 0,\ z_0 \geq 0, \quad q + dz_0 + Mz \geq 0$$

with the property:

$$q^T z + z^T M z = 0 \quad \text{and} \quad z_0 = 0.$$

Let $C(q, M)$ and $C(d, q, M)$ denote the solution sets of problems q/M and $d/q/M$, respectively. Of particular interest in the following will be the special LCP d/M and its solution set $C(d, M)$.

We will use the following definitions and results, which require no elaboration.

Definitions of two matrix classes as those matrices satisfying:

$E(d)$: $z \in C(d, M)$, $z \neq 0 \implies \exists\, \hat{z} \geq 0$, $\hat{z} \neq 0$ such that:

(i) $-M^{\mathsf{T}}\hat{z} \geq 0$

(ii) $z \geq \hat{z}$ and $d + Mz + M^{\mathsf{T}}\hat{z} \geq 0$.

$E^*(d)$: $z \in C(d, M) \implies z = 0$.

Lemma 3.1. $E(d) = E^*(d)$ for any $d > 0$ or $d < 0$.

Observation. $E(0) = L_2$ of Eaves [Eav 71].

Corollary 5.2. Let $d/q/M$ be partitioned as:

$$\begin{bmatrix} d_J \\ d_K \\ 0 \end{bmatrix} \Big/ \begin{bmatrix} q_J \\ q_K \\ r \end{bmatrix} \Big/ \begin{bmatrix} M_{JJ} & M_{JK} & b \\ M_{KJ} & M_{KK} & 0 \\ -a^{\mathsf{T}} & 0 & 0 \end{bmatrix},$$

where M_{JJ} and M_{KK} are square matrices, a and b are columns, d_J, d_K, r, a and b are all positive, $M \in E(0)$, and $M_{KK} \in E(d_K)$.
If $C(d, q, M)$ has a secondary ray, then q/M is infeasible.

Commentary

If the partition K is vacuous, then the structure and results of the above corollary correspond to those of Theorem (11.5) of Eaves [Eav 71].

It is important to note that the covering vector d is specially constructed to be strictly positive except for the zero in the last position. Since this is not a standard specification, in practice it would be necessary to modify the default specification of the covering vector incorporated in available software for implementing Lemke's algorithm.

6.2 Application to a linearized subproblem

We are concerned with the linearized subproblem encountered at any iteration of any of the sequential methods we study. For simplicity, we will suppress the superscripts referring to iteration number. The results also do not depend on the nature of the linearization employed, so we will non-specifically refer to the linearized demand functions as $b + Bp$. Where we depart from previous modes of presentation is that now we take explicit account of the partitioning of commodities into final commodities and primary or intermediate commodities.

Accordingly, partition the vector of prices as $p = (\pi, \sigma)$, where π corresponds to final commodities and σ to primary or intermediate commodities. We also, with a suggestive abuse of notation, use π and σ to denote index sets for partitioning the rows of the activity analysis matrix A, the elements of the vector b, and the rows and columns of the matrix B. Since the σ partition corresponds to commodities for which demand is constant, it follows that $B_{\sigma\sigma}$ would be zero in any sensible linearization. $B_{\sigma\pi}$ may also be zero, as in the case $B = \nabla d(\pi, \sigma)$. $B_{\sigma\pi}$ would not be zero, for instance, if $B = \left[\nabla d(\pi, \sigma) + \nabla^{\mathsf{T}} d(\pi, \sigma)\right]$. If $B_{\sigma\pi} = 0$, b_σ is constant across all iterations and represents the negative of aggregate endowments (which by definition are zero for intermediate commodities).

With this notation, we may state the linearized subproblem as:

Find a complementary solution of:

$$
\begin{array}{lrcrcrcrcl}
\text{(SP1)} & -B_{\pi\pi}\pi & - & B_{\pi\sigma}\sigma & + & A_{\pi\cdot}y & + & hv & \geq & b_\pi \quad \perp \quad \pi \geq 0 \\
\text{(SP2)} & -B_{\sigma\pi}\pi & & & + & A_{\sigma\cdot}y & & & \geq & b_\sigma \quad \perp \quad \sigma \geq 0 \\
\text{(SP3)} & -(A_{\pi\cdot})^{\mathsf{T}}\pi & - & (A_{\sigma\cdot})^{\mathsf{T}}\sigma & & & & & \geq & 0 \quad \perp \quad y \geq 0 \\
\text{(SP4)} & -h^{\mathsf{T}}\pi & & & & & & & = & -1 \quad \perp \quad v \\
\text{(SP5)} & \pi & , & \sigma & , & y & & & \geq & 0
\end{array}
$$

In order to correspond exactly to the structure of Garcia, we must now write the normalization equality (SP4) as two inequalities and divide the artificial variable v into its positive and negative parts, denoted by v_+ and v_-, respectively. Given these notational conventions, we can partition the data of the linearized subproblem so as to correspond directly to the partitioning of Garcia's Corollary 5.2 (above):

$$
\left[\begin{array}{c} d_\pi \\ \hline d_\sigma \\ d_y \\ 1 \\ \hline 0 \end{array}\right] \Big/ \left[\begin{array}{c} -b_\pi \\ \hline -b_\sigma \\ 0 \\ -1 \\ \hline 1 \end{array}\right] \Big/ \left[\begin{array}{c|cccc|c} -B_{\pi\pi} & -B_{\sigma\sigma} & A_{\pi\cdot} & -h & +h \\ \hline & & A_{\sigma\cdot} & & \\ -(A_{\pi\cdot})^{\mathsf{T}} & -(A_{\sigma\cdot})^{\mathsf{T}} & & & \\ -h^{\mathsf{T}} & & & & \\ \hline -h^{\mathsf{T}} & & & & \end{array}\right].
$$

Before verifying that this data satisfies the hypotheses of the corollary, it is worth mentioning why it is advantageous to restrict the price normalization to the final commodity prices alone. As a theoretical issue, it does not matter. As a computational issue, excluding the prices σ from the normalization leaves a problem structure that requires each subproblem to be feasible with respect to the commodity balances (SP2), which are identical to the original problem constraints and do not change from iteration to iteration. If the artificial column extended into these constraints, then feasibility would be attained only at equilibrium where $v = 0$. We may thus expect a more reasonable and constrained path to the solution if the artificial column is limited to the final commodity constraints. If $B_{\sigma\pi} = 0$, we also may be

able to take computational advantage of any simple bounds that appear in the constraints (SP2). This also could not be done if the artificial column extended into these constraints.

We now verify that the problem data for any linearized subproblem will satisfy the conditions for Garcia's corollary if $h > 0$. Note that h corresponds to both a and b of the matrix in the corollary, which are required to be positive. We may presume that all components of the covering vector (except the final zero) are positive as required.

Claim: $M_{KK} \in E^*(d_K)$.
Proof. M_{KK} is clearly skew-symmetric. Hence, $z_K^T M_{KK} z_K = 0$ for all z_K. Since $d_K > 0$, if $z_K \geq 0$ and $z_K^T d_K + z_K^T M_{KK} z_K = 0$, it must be that $z_K = 0$. That is, $z_K \in C(d_K, M_{KK}) \implies z_K = 0$.

Claim: If $h > 0$, then $M \in E(0)$.
Proof. Let $z = (\pi, \sigma, y, v_-, v_+) \in C(d, M)$ and $\hat{z} = (\hat{\pi}, \hat{\sigma}, \hat{y}, \hat{v}_-, \hat{v}_+)$ be the two vectors relevant to the definition of class $E(d)$ above. Here, we are temporarily concerned with $d = 0$. Given that $h > 0$, it is clear that $z \geq 0$ and $Mz \geq 0$ requires $\pi = 0$. Similarly, $\hat{z} \geq 0$ and $-M^T\hat{z} \geq 0$ requires $\hat{\pi} = 0$. (As an interesting aside, note that the submatrix of M obtained by deleting the rows and columns corresponding to π is skew-symmetric. Thus, for any z satisfying $z \geq 0$ and $Mz \geq 0$ we have $z^T M z = 0$, i.e., $z \in C(0, M)$.) Given that $\pi = \hat{\pi} = 0$, we can write:

$$Mz + M^T\hat{z} = \begin{pmatrix} -B_{\sigma\sigma}\sigma & +A_{\pi\bullet}(y - \hat{y}) - h(v_- - \hat{v}_-) + h(v_+ - \hat{v}_+) \\ & A_{\sigma\bullet}(y - \hat{y}) \\ -(A_{\sigma\bullet})^T(\sigma - \hat{\sigma}) & \\ & 0 \\ & 0 \end{pmatrix}.$$

We now consider two cases. First, if $\sigma = 0$, then we simply take $\hat{z} = z = (0, 0, y, v_-, v_+)$. Then $-M^T\hat{z} = Mz \geq 0$ and $Mz + M^T\hat{z} = 0$. Second, if $\sigma \neq 0$ we take $\hat{z} = (0, \sigma, 0, 0, 0) \leq z$. Then

$$-M^T\hat{z} = \begin{pmatrix} 0 \\ 0 \\ -(A_{\sigma\bullet})^T\sigma \\ 0 \\ 0 \end{pmatrix} \geq 0 \quad \text{and} \quad Mz + M^T\hat{z} = \begin{pmatrix} -B_{\sigma\sigma}\sigma + A_{\pi\bullet}y - hv_- + hv_+ \\ A_{\sigma\bullet}y \\ 0 \\ 0 \\ 0 \end{pmatrix} \geq 0.$$

So both cases satisfy the conditions (i) and (ii) of the definition of $E(0)$. Hence, $M \in E(0)$ provided $h > 0$.

Since the problem data satisfy both of the conditions for Garcia's Corollary 5.2, we can conclude that each linearized subproblem is in a class that can be solved by Lemke's method provided that the problem is feasible. But if $h > 0$ the only way that the subproblem can be infeasible is for the inequalities (SP2)-(SP5) to be inconsistent. These inequalities do not change from subproblem to subproblem, however, and their inconsistency would imply that no equilibrium solution exists. Since any properly formulated general equilibrium model does have an equilibrium solution, we can conclude (as a theoretical matter) that each linearized subproblem has a complementary solution which can be found by Lemke's method.

7

COMPUTATIONAL COMPARISONS

In this chapter we present the results of some computational experiments applying five different solution methods to each of two small test problems. The two problems are reasonably well known in the equilibrium literature and are small enough to permit hundreds of model solutions from different starting points. Despite the small size, they possess properties which serve to illustrate much of what can go right and wrong in solving a general equilibrium model. The solution methods implemented are as follows:

1. SLCP (Section 4.1) applied to problem Eq-NLCP (Section 3.2)
2. SLP (Section 4.4) applied to problem Eq-NLP (Section 3.3)
3. symmetric positive semidefinite linearization (Section 4.6) applied to Eq-NLCP
4. projected (augmented) Lagrangian method (Section 4.5) applied to Eq-NLP
5. direct solution of problem Eq-DEPF (Section 5.2.1).

Another linearization strategy discussed in Section 4.6 was abandoned after disappointing intitial results. In practice it turns out that the skew-symmetric approximation matrix $\left[\nabla d(p) - \nabla^{\mathsf{T}} d(p)\right]$ is prone to be so rank-deficient as to prevent the attainment of a basic solution of the linearized subproblem that has all positive prices. As all equilibrium price vectors for the two test problems are strictly positive, a method based on pivotal solutions of the subproblems could never find an equilibrium using this linearization.

If there is a single word to describe the results obtained, it is diversity. A few generalizations are supported, but there are numerous exceptions to every conclusion. Before presenting the results, we briefly describe the test problems and discuss the specific implementations of the solution methods.

7.1 Two test problems

The first problem was devised by Scarf [Sca 73, pp. 113–119] and has become a standard problem for testing and comparing solution methods for general equilibrium problems. The model contains 14 commodities and 26 linear production activities. The commodities may be subdivided into 7 final consumption commodities (for which there are no initial endowments), 3 primary commodities, and 4 intermediate commodities (these terms were defined in Section 2.1). Aggregate final demand is the sum of the demands of 4 consumers, each with a Cobb-Douglas utility function. This means that demand for final commodity i becomes unbounded as $p_i \to 0$. If none of the primary commodities has a positive price, then income is zero and so is demand. The Scarf model has a unique equilibrium.

In [Keh 85] Kehoe describes a tiny equilibrium problem specifically constructed so as to have multiple equilibria. The model has 4 final commodities and 2 linear production activities.

Final demand is the sum of the demands of 4 Cobb-Douglas consumers, each with an initial endowment of only a single commodity. The problem was devised so as to have an unstable equilibrium at unit prices, which implies the existence of at least two other equilibria. (Proof of this relies upon the global index theorem described in both [Keh 82] and [Keh 85].) Kehoe establishes that the model has exactly 3 equilibria. When prices are normalized to sum to 4, these are:

$$\bar{p} \in \begin{cases} (0.6377,\ 1.0000,\ 0.1546,\ 2.2077) & \text{(Eq1)} \\ (1.0000,\ 1.0000,\ 1.0000,\ 1.0000) & \text{(Eq2)} \\ (1.1005,\ 1.0000,\ 1.2346,\ 0.6649) & \text{(Eq3).} \end{cases}$$

Both production activities are strictly positive at all equilibria, which results in p_2 having the same value for all equilibria. This is because a binding excess profit constraint for the second production activity requires that $p_1 + p_3 + p_4 = 3p_2$. The intersection of this and the price normalization fixes $p_2 = 1$, which in turns restricts the values of the other three prices to lie on the subsimplex $p_1 + p_3 + p_4 = 3$. The other binding profit constraint then forces all equilibria to lie on a line in this subsimplex. Indeed, any iterates obtained in solving the model must lie on this line if both production activities are positive. We found these observations in unpublished work of Mathiesen and Rutherford [MR 83], who in turn attribute them to unpublished work of Kehoe (which later appeared in [Keh 84]). These special features of the model prove to have significant consequences for the computational performance of the methods studied.

7.2 Solution methods implemented

All five methods are implemented by means of specialized user subroutines in MINOS. In particular, SLCP, SLP, and the symmetric linearization method are all implemented by means of the same program to sequentially revise the linearized constraints and define the appropriate objective function for the subproblem. MINOS is used solely to manage and solve the individual subproblems, not to "supervise" the overall solution process. In contrast, the other two methods directly apply MINOS as an optimization algorithm for solving problems Eq-NLP and Eq-DEPF. Use of a common computational vehicle makes the experiments more manageable and also enhances the comparability of the solution times for the various methods.

7.2.1 Departures from standard SLCP method

Using MINOS to solve the subproblems means that SLCP is actually implemented as (Wilson's) SQP method (Section 4.2) applied to problem Eq-NLP. Thus, the LCP subproblems are *not* processed by Lemke's method, but rather are solved as indefinite quadratic programs. With this approach, it is possible that some subproblems will not be solved to global optimality (i.e., complementarity). This in fact happens from many starting points at one or two of the early iterations of the method. Nonetheless, failure to achieve complementarity

does not prevent convergence for either problem from any starting point. The consequence of the MINOS implementation is that the "simulated" SLCP iterations follow a somewhat different path than would a "pure" implementation.

Another difference in implementation arises from the choice of step length and its relation to avoiding zero prices at linearization points. The most elegant approach would be to implement a rigorous line-search procedure, applying a merit function (such as that of Eq-DEPF) that becomes infinite at points outside the domain of the demand functions. We regret that time did not permit undertaking the programming effort required to accomplish such a solution. (We do, however, evaluate the merit function at each point obtained by a sequential method and therefore can assess progress after the fact.) Any other appoach to determining a step length and avoiding boundaries is necessarily heuristic. Mathiesen's approach is to allow zero prices in the LCP solutions but then to take less than a unit step if one or more of the prices is zero. This means that the merit function cannot be evaluated at the LCP solution if there is a zero price; it must be evaluated at the linearization point determined by the step length. A more attractive and convenient alternative in the context of our implementation with MINOS is simply to impose reasonable nonzero lower bounds on the prices. This means that the merit function can be evaluated at all subproblem solutions and that a unit step is always feasible. With this approach, however, a binding lower bound in the subproblem can and generally does prohibit the attainment of a complementary solution. For the most part, these bounds are the true cause for the lack of complementarity in certain early iterations that we alluded to above.

Intuitively, we know that obtaining a zero price in a linearized subproblem means that the linear approximation did not properly reflect the true boundary behavior of the demand functions. It is thus arguable that obtaining such a solution is of no more value than obtaining a non-complementary solution with more reasonable prices. In some sense, we might expect that preventing the early iterations from obtaining extreme prices would enhance overall convergence. As a crude test of this hypothesis, we applied the first three methods to each problem using lower bounds of 0.01 and 0.0001. On balance, use of the 0.01 bounds proved superior, so the rest of the results pertain to computations using the higher bounds.

7.2.2 A symmetric linearization

For the third method we define a sequential linearization scheme that utilizes a symmetric, negative semidefinite approximation matrix (referred to as B^k in Section 4.6). Since the demand functions enter as $-d(p)$ in the commodity balances, it is $-B^k$ that appears in the linearized constraints. The resulting constraint matrix is then bisymmetric and positive semidefinite, implying that the subproblem can be solved as a convex QP on the price space alone. The production activity levels are obtained from the dual, i.e., from the Lagrange multipliers on the excess profit constraints. We derive this approximation matrix from the sums of the so-called Slutzky matrices for the demand functions of the individual consumers. For a consumer with income θ and demand function $d(p,\theta)$, the Slutzky matrix $S(p,\theta)$ is

defined as:

$$S(p,\theta) \equiv \nabla_p d(p,\theta) + \nabla_\theta d(p,\theta)\,[d(p,\theta)]^T.$$

(The derivation may be found in any intermediate textbook of microeconomics.) The matrix $S(p,\theta)$ is symmetric and negative semidefinite; in particular, p is in its null space. These properties are clearly preserved under summation across a finite collection of consumers. In both of our test problems, we know the demand of each consumer as a function of prices and income, as well as the definition of individual income in terms of endowments and prices. Hence, at any price vector p we can construct the individual Slutzky matrices and obtain their sum. We thus obtain a meaningful symmetric, negative semidefinite linearization of the aggregate demand function $d(p)$.

Note that the Slutzky matrix is properly defined only for final consumption items. It is thus significant that, in a problem with primary and/or intermediate commodities (such as the Scarf problem), an approximation matrix derived from Slutzky matrices is nonzero only in the submatrix corresponding to prices of the final commodities. Such an approximation not only has $B_{\sigma\pi} = 0$ in the rows designated (SP2) in Section 6.2, it further has $B_{\pi\sigma} = 0$ in inequalities (SP1). Other symmetric approximations, such as any scheme involving $\left[\nabla d(\pi,\sigma) + \nabla^T d(\pi,\sigma)\right]$, produce a nonzero $B_{\sigma\pi}$ in inequalities (SP2). These terms not only increase the density of the matrix, they also produce a distortion in the commodity balances for the primary and intermediate commodities.

We are not aware of any previous attempts to use Slutzky matrices as a basis for linearizing final demand functions. It is unfortunate that such a construction is not possible for an arbitrary nonintegrable aggregate demand function. It is applicable only to problems involving sums of known individual demand functions for which the Slutzky matrices can be constructed. In the following discussions we will use the abbreviation SLTZ to refer to this sequential linearization method based on Slutzky matrices.

7.2.3 Using MINOS to solve Eq-NLP and Eq-DEPF

As described in Section 4.5, the features that characterize the projected augmented Lagrangian method (as implemented in MINOS to solve nonlinearly constrained problems) are the Lagrangian term based on multiplier estimates and the penalty parameter weighting the quadratic penalty term. If both terms are suppressed, the resulting procedure automatically implements SLP (without trust regions) on any problem with a linear objective function (such as Eq-NLP). In our context, this simple, low-overhead implementation of SLP naturally proves to be somewhat faster than the specialized user program we devised to implement all three of methods SLCP, SLP, and SLTZ. For maximum comparability, we will report the results for these three methods based on common use of the special user program.

The interesting question with respect to using MINOS on Eq-NLP is whether the additional incorporation of the Lagrangian and/or penalty terms proves to be a help or a hindrance relative to SLP. The terms impart to each subproblem some of the underlying nonlinear nature of the original problem, and the penalty term in particular is intended to provide

a stabilizing influence that promotes global convergence. This benefit comes at the cost of solving subproblems with a general nonlinear objective, which is furthermore nonconvex in the case of Eq-NLP. To address this question, we solved each test problem from many different starting points, both with and without the Lagrangian term and for different values of the penalty parameter.

An important aspect of solving problem Eq-DEPF directly is that no initial values for the prices need to be supplied. A standard Phase I procedure will automatically generate a feasible solution of the linear constraints; see again Section 5.2.1. For this solution, x, y and v may very well be zero, but the prices determined will be normalized and feasible with respect to the excess profit constraints. Thus, the first evaluation of the demand functions occurs at feasible prices. This is in sharp contrast to all the other solution methods which require the user to divine a set of starting prices. These may or may not be feasible depending on the method used to obtain them. As it is meaningless to attempt to solve Eq-DEPF from numerous starting points, the only relevant comparison is between a single solution time for Eq-DEPF and the distribution of solution times for the other methods. The interesting issue in solving Eq-DEPF is the effect on solution time of differing values of the penalty parameter in the objective function. This is readily investigated, and we report below the effects of this parameter on solution times for each test problem.

7.2.4 Specification of starting points

Since our purpose is to investigate both the robustness and speed of convergence of the various methods from arbitrary starting points, we deem it preferable to avoid randomization and instead construct a uniform grid of prices covering the (interior of the) relevant price simplex. For the Scarf problem, we construct a grid over the 10 prices for final and primary commodities. To obtain a manageable number of points, the range for each price is subdivided into 3 (equal) intervals, the endpoints of which define a grid of 220 points on the simplex. For the Kehoe problem, we can afford a finer grid based on 10 intervals, which results in 286 starting points. For each problem, the minimum allowable starting price is 0.01. This is relative to price normalizations requiring the 4 prices in the Kehoe model to sum to 4 and the 7 final commodity prices in the Scarf model to sum to 7.

It is important to bear in mind that the purpose of this grid of prices is to examine how well the methods perform even when given a starting point near the boundary of the simplex — and hence near the boundary of the domain of definition of the demand functions $d(p)$. Many of the points have one or more prices at the minimum of 0.01. Indeed, for the Scarf problem each starting point has at least 7 of the 10 prices at the minimum value. For the Kehoe problem, 202 of the 286 points have at least 1 of the 4 prices at the minimum value. Such points naturally tend to produce rather unstable initial iterations. This allows for some worst-case comparisons, but such results are not necessarily representative of the performance of the method in a realistic setting.

In practice, model users may well have reasonable "priors" as to relative prices and, at the very least, can readily avoid specifying initial prices near the boundary of the domain of the demand functions. To assess the performance of the methods when given more reasonable

starting points, we do the following. For the Scarf problem, we generate another 220 starting points by taking a convex combination of the near-boundary points and the center of the simplex (i.e, a vector of ones). We use a weight of 0.7 on the center point. (For reference, all but one of the equilibrium prices are near unity; the "outlier" is approximately 0.5.) In contrast, for the Kehoe problem, we simply examine the results for the subset of 84 starting points that are not near the boundary.

7.2.5 Convergence criteria

The significant feature of imposing lower bounds on the prices is that any subproblem solution for methods SLCP, SLP, and SLTZ produces a feasible solution for problem Eq-DEPF. We can then utilize the value of the Eq-DEPF objective function as a rigorous measure of progress for the method and also as a viable stopping criterion. For the convergence criterion, we established the extremely strict requirement that both the value of the Eq-DEPF objective and the absolute value of artificial variable v must be less than 10^{-7}. This was chosen after early experiments indicated inadequately precise convergence of the SLTZ method with a criterion of 10^{-6}, which is the tolerance for judging feasibility and optimality in MINOS. These MINOS tolerances were also tightened to 10^{-7} for the SLTZ method. The stoppping criterion utilized is much more demanding than the kinds of criteria typically employed in solving equilibrium models.

In using MINOS to solve problem Eq-DEPF, we also found it worthwhile to use the tighter 10^{-7} feasibility and optimality tolerances. This did not appreciably increase solution time, but did improve the accuracy of the final solution. For solving problem Eq-NLP with MINOS, the default MINOS criteria for convergence proved more than adequate to ensure convergence within the tolerances defined above for the other methods.

7.2.6 Hardware, software and timing measures

All of the computational experiments were performed on an IBM 3090 Series 200 computer, kindly made available to us by the IBM Palo Alto Scientific Center. The operating system is VM/CMS running as a "guest host" under VM/XA. Both MINOS and the required user subprograms were compiled and executed using Release 2 of Version 2 of VS FORTRAN. No attempt was made to utilize the available vectorization and parallel processor capabilities. Even without using such features, 3090 computing power is something of an "overkill" for the problems solved. As a result, we will be comparing solution times measured in fractions of a second. For the purposes of our comparison of methods, it is the relative solution times that matter. All reported solution times exclude program linking, MINOS initialization, and printing of the final solution.

7.3 Results for the Scarf problem

The existence of a unique equilibrium solution for the Scarf problem allows for a straightforward assessment of computation times. It is meaningful to examine the range and average

of solution times across all starting points as well as ratios of solution times for different methods from the same starting points. To begin with the simpler discussions, we first address using MINOS to solve problems Eq-DEPF and Eq-NLP. We then examine at greater length the comparisons of methods SLCP, SLP, and SLTZ.

7.3.1 Solving Eq-DEPF with MINOS

As we indicated above, solving problem Eq-DEPF does not require the specification of an initial starting point. What we examine is the effect of the choice of penalty parameter on solution time. We solved the Scarf model a small number of times for each of 5 different settings of the penalty parameter. After averaging out some small random fluctuations in the individual trials, representative solution times are as follows:

parameter	0.01	0.1	10	10	100
CPU seconds	0.145	0.132	0.145	0.140	0.149

The relative variation is not particularly significant, and we have no explanation for the bimodal pattern. A solution time on the order of 0.14 seconds for Eq-DEPF should be kept in mind when reviewing the results for the other methods below.

7.3.2 Solving Eq-NLP with MINOS

We solve problem Eq-NLP using four different strategies:

	Lagrangian term	penalty parameter
(1)	no	0
(2)	yes	0
(3)	no	10
(4)	yes	10

Strategy (1) is effectively SLP because of the linearity of the objective function of Eq-NLP. Since no Lagrange multiplier estimates are available for the first linearized subproblem, strategies (1) and (2) solve the same initial subproblem, and so do strategies (3) and (4).

The most significant result of these comparisons is that all four strategies succeed in locating the equilibrium solution from all 220 near-boundary starting points. Solution times range from 0.1 to 0.5 seconds across all starting points. We can also compare solution times across methods initiated from the same starting points. We do so by computing the ratios of the solution times for strategies (2)–(4) to that of strategy (1). With only two exceptions, the solution times for strategies (3) and (4) range between 95% and 102% of that for strategy (1); the average is about 99%. Curiously, the number of linearizations required from each starting point is the same for all three of these strategies. Hence, the presence of a penalty

term did not affect the convergence rate obtained, and, contrary to prior expectations, the more complicated objective function for the subproblem did not increase overall solution time. On balance, we do not consider the variation in solution times for these three strategies to be at all significant.

In contrast, the solution times for strategy (2) range from 62% to 153% of that for strategy (1); the average is about 105%. In most, but not all, cases the number of linearizations was reduced by the presence of the Lagrangian term. Nonetheless, solution times were frequently longer despite the improved convergence rate. There is no apparent pattern to explain why solution times increased for some starting points and decreased for others. In particular, the distance of the starting point from the equilibrium point is not in any way related to the solution time. We conjecture that the wide variation in relative solution times is attributable to differing qualities of the Lagrange multiplier estimates obtained from the first two or three subproblems. In the absence of a stabilizing penalty term, the multiplier estimates from the first subproblem are likely to be quite poor, given a linearization point near the boundary of the simplex. Poor multiplier estimates can in turn lead to distorted solutions for the next subproblem, which may well take longer to compute if the solution is distant from that of the previous subproblem solution.

Since the primary purpose of these comparisons is to assess the necessity of the Lagrangian and penalty terms for achieving global convergence from near-boundary starting points, we do not think it worthwhile to repeat the experiment from the interior starting points.

7.3.3 Three sequential methods

We first discuss the results for the set of starting points with at least 7 of the 10 prices set at the minimum value of 0.01. It is significant that all three of the methods SLCP, SLP, and SLTZ succeeded in finding the equilibrium solution from all 220 of these deliberately poor starting points. The character of the results partition the starting points into two groups. The most varied results are obtained for a group of 84 points at which all 3 of the prices for primary commodities are set at 0.01. (This means that consumer income is near zero, since the Scarf model has no endowments of the final consumption goods.) We use the solution times for SLCP as a basis for comparison. These range from 0.19 to 0.46 seconds across all 84 points. For the most part, the number of linearizations required is in the range of 4–6, but 10 troublesome points required 10–12 linearizations. The variation is not related in any visible way to the distance of the starting point from the equilibrium.

We now look at solution times (from the same starting point) for the SLP and SLTZ methods expressed relative to the solution time for SLCP. For SLP these time ratios range between 0.37 and 1.54, with an average of 0.87. The most extreme of these deviations are attributable to significantly different numbers of linearizations required. For the most part, but not universally, solution times for the same number of linearizations are markedly lower for SLP. For SLTZ, the range of relative solution times is from 0.14 to 0.78, with an average of 0.48. In a small number of cases, SLTZ required fewer linearizations than SLCP, but it generally required 2 or 3 more — and occasionally 2 or 3 times as many. These excursions reflect an underlying phenomenon that SLTZ iterations occasionally produce an increase in

the value of the Eq-DEPF merit function. These uphill segments of the path may endure for a few linearizations, but descent is eventually regained. Lower overall solution times in the face of the higher linearization counts clearly point to substantially lower solution times for solving the smaller SLTZ subproblems.

The results are quite different and less diverse for the other 136 points. The solution times for SLCP range from 0.41 to 0.88 seconds, which is almost wholly above the range of solution times for the other group of points. Here, the number of linearizations required is generally in the range 9–12. The *relative* solution times for SLP are more tightly confined in the range 0.32–0.71, with an average of 0.45. For the most part, the number of linearizations is somewhat higher than that for SLCP, so these time savings reflect lower solution times for the individual subproblems. Quite surprisingly, on this group of points SLTZ generally required fewer linearizations than either SLCP or SLP. We do not know how to account for this result, but it naturally means that solution times are substantially lower for SLTZ. Solution times relative to SLCP are in the range 0.14 to 0.35, with an average of 0.21.

The above results certainly attest to the robustness of all three sequential methods when applied to a well-behaved problem such as the Scarf model. We now look at more "normal" performance as measured by the other set of 220 points, for which no starting price is lower than 0.7. Here, we tabulate the range and average solution times for the three methods so as to permit a somewhat meaningful comparison with the time required to solve problem Eq-DEPF directly.

CPU seconds	SLCP	SLP	SLTZ
minimum	0.157	0.125	0.018
maximum	0.393	0.285	0.075
average	0.219	0.170	0.034

An Eq-DEPF solution time of 0.14 seconds is basically at the lower end of the range of solution times for SLCP and SLP, but it is almost twice the value of the *upper* end of the range for SLTZ.

Turning to the relative solution times from identical starting points, the results show a strong concentration near the average with a significant number of outliers above the average. SLP solution times relative to SLCP range between 0.59 and 1.68, with a mean of 0.79. For SLTZ the range is 0.1–0.36, with an average of 0.16. Underlying causes for the deviations are essentially the same as those discussed above for the collection of points with near-zero endowment prices. No pattern appears with respect to the distance of the starting point from the equilibrium, but there is an interesting pattern with respect to the solution time for SLCP. This pattern is discernible in Figures 1 and 2, which plot the relative solution times of SLP and SLTZ versus the solution time for SLCP. (The highest outlier was excluded from each figure so as to obtain better resolution for the rest of the points.) The general downward trend in the clusters of points indicates that, on groups of starting points for which SLCP encountered increasing difficulty, the other two methods suffered less than proportional increases in solution time. The scatter also indicates that SLP and SLTZ

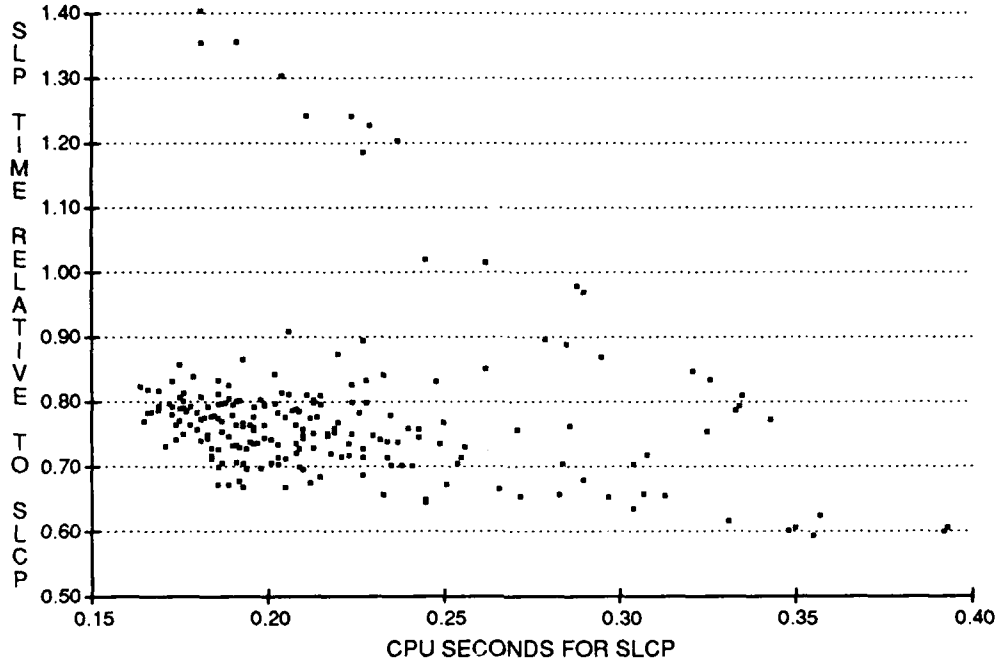

Figure 1: SLP vs. SLCP — Scarf problem

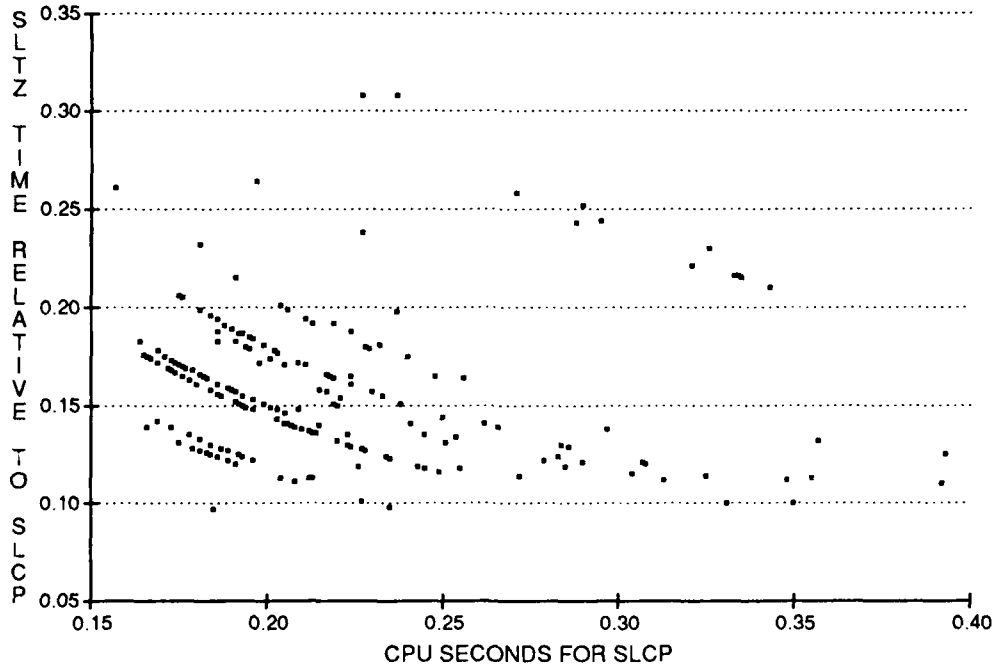

Figure 2: SLTZ vs. SLCP — Scarf problem

frequently, but not universally, encountered difficulties with the same starting points. The conclusions which state themselves, however, are the overwhelming superiority of the SLTZ method on the Scarf problem and the generally superior performance of SLP relative to SLCP.

7.4 Results for the Kehoe problem

The existence of three distinct equilibrium solutions for the Kehoe problem rather complicates the assessment of computational results. Examining the range and average of solution times for a given method across all starting points can provide an indication of how long it takes the method to find *some* equilibrium point, but both range and average may depend upon the proportion of times each equilibrium point is found. Also, it does not seem particularly meaningful to compare solution times for different methods from the same starting point unless the methods converge to the same equilibrium point. It is interesting to note when different methods converge to different points, but it is not clear what else can be said about such cases. To begin again with the simpler discussions, we first address using MINOS to solve problems Eq-DEPF and Eq-NLP. We then examine at greater length the comparisons of methods SLCP, SLP, and SLTZ.

7.4.1 Solving Eq-DEPF with MINOS

As for the problem of Scarf, we solve the Eq-DEPF formulation of the Kehoe problem for 5 different settings of the penalty parameter. For all parameter values, the solution finds the (Eq1) equilibrium point. Again, the variation is not at all significant: 0.018 seconds is required for a parameter value of 0.1, with the other values requiring 0.02 seconds. Convergence to (Eq1) with a solution time on the order of 0.02 seconds for Eq-DEPF should be kept in mind when reviewing the results for the other methods below.

7.4.2 Solving Eq-NLP with MINOS

We solve the Eq-NLP formulation of the Kehoe problem using the same four strategies as identified above for the Scarf problem. Again, the most significant result is that, from all 286 starting points, all four strategies succeed in finding one of the equilibrium solutions, including the unstable equilibrium on several occasions. It is interesting that, from each starting point, strategies (2)–(4) converge as a group to the same equilibrium point, which almost half the time is a different point than that found by strategy (1). Moreover, with only one exceptional point, the solution times and number of linearizations for strategies (2)–(4) are virtually identical.

For all strategies, solution times range from about 0.02–0.11 seconds across starting points. Great variation is apparent in the relative solution times of similarly behaving strategies (2)–(4) as compared to strategy (1). For the 115 points from which all strategies converge to (Eq1), relative solution times range from 51% to 265%, with an average of 138%. Only 9 points result in all strategies converging to the unstable equilibrium; here the range is much

tighter (88% to 107%) and the mean is 103%. Finally, all strategies converge to (Eq3) from 37 points. The range in relative solution times is 61% to 187%, with an average of 117%. Again, there is no apparent pattern to explain these variations in relative performance. Moreover, the results for the subset of 84 interior starting points very much resemble the results for the entire set. With such a diversity of results, we are in no position to draw any viable conclusions or generalizations.

7.4.3 Three sequential methods

As we indicated earlier, of the 286 starting points generated for the Kehoe problem, 202 points (intentionally) have at least one price at the minimum of 0.01. While we might intuitively expect to see markedly different behaviors of the methods in the subset of near-boundary points as compared to the subset of 84 interior points, this did not turn out to be the case for the Kehoe problem. In light of this result, we will discuss all 286 cases together in order to utilize the larger sample size.

Both SLCP and SLP successfully converged to an equilibrium from all starting points. SLCP never reached the unstable equilibrium (Eq2), while SLP converged to (Eq2) on 14 occasions. SLTZ also never found the unstable equilibrium and converged to (Eq1) from 181 starting points. From the other 105 points, SLTZ clearly targeted (Eq3) as a limit point, but the convergence criterion was met before the iterates obtained a precision comparable to that of the other methods. For instance, the terminal values for p_1 were either 1.1003 or 1.1008 (depending upon the direction of approach). This is representative of a general pattern for SLTZ on the Kehoe model that iterations terminate upon narrowly satisfying the convergence criterion. In contrast, the final iteration of SLCP or SLP almost always satisfies conditions several orders of magnitude stronger than the specified convergence criterion. This precise "jump" at the last iteration is typical of Newton's method. This difference in convergence patterns is very important, but we do not believe it justifies declaring that the SLTZ method failed to converge; it just fails to converge at a satisfactory rate (on the Kehoe problem).

It is significant that, out of 286 starting points, all three methods converged to (Eq1) in only 71 cases and to (Eq3) on only 29 occasions. Pairwise intersections are also important in terms of defining subsets of starting points over which to evaluate relative solution times. SLCP and SLP jointly converge to (Eq1) from 83 starting points and to (Eq3) from 35 points. SLCP and SLTZ jointly converge to (Eq1) from 146 starting points and to (Eq3) from 71 points.

Before presenting results on relative solution times, we again tabulate the range and average solution times for the three methods across all starting points. Given different patterns as to which equilibrium is located, these figures are not as meaningful as those for the Scarf problem, but they do allow a loose comparison with the time required to solve problem Eq-DEPF directly.

CPU seconds	SLCP	SLP	SLTZ
minimum	0.025	0.016	0.086
maximum	0.111	0.097	1.382
average	0.077	0.057	0.475

An Eq-DEPF solution time of 0.02 seconds is basically at the lower end of the range of solution times for SLCP and SLP. The times for SLTZ lie almost wholly above the upper end of the ranges for SLCP and SLP. This is in strong contrast to the results for the Scarf problem.

We examine relative solution times from identical starting points only for points from which the methods compared converged to the same equilibrium. As for the Scarf problem, the results show strong concentrations near the averages with significant numbers of outliers above the averages. SLP solution times relative to SLCP range between 0.38 and 1.89, with a mean of 0.75. This is quite similar to the results for the Scarf problem. For SLTZ, the story is completely different: the range is 1.4–32, with an average of 6.1. No pattern appears with respect to the distance of the starting point from an equilibrium. Indeed, as can be seen from Figure 3, there is no particular pattern at all to the relative performance of SLP. In sharp contrast, Figure 4 illustrates that the generally poor performance of SLTZ is markedly worse when the process converges to (Eq3). Note that the two different symbols for points in the figures correspond to the two different equilibrium points obtained. We again have excluded the highest outlier from each figure.

The decidedly poor performance of SLTZ on the Kehoe problem is attributable to the following general pattern of convergence. The worst cases arise when the first subproblem solution that obtains positive levels for both production activities also obtains prices that are near the unstable equilibrium (Eq2). The iterates never converge to this equilibrium — no matter how close it may be — but proceed toward another equilibrium at a linear rate. Along this path, the Eq-DEPF merit function increases for many iterations, often by several orders of magnitude. A peak is eventually reached, and descent is resumed at a very poor linear rate. It is remarkable that convergence is achieved at all. Intuitively, this pattern suggests that the unstable equilibrium exerts a perverse kind of attraction which does not draw the iterates toward (Eq2) but seriously retards the rate at which the iterates are drawn to one of the other equilibria. The proximity of (Eq3) to (Eq2) causes this effect to be particularly pronounced for iterates that target (Eq3).

7.5 Commentary

The great diversity of the above results makes it difficult to derive any meaningful generalities and conclusions. It is important to recognize, however, that all of the methods studied did in fact converge to an equilibrium, regardless of the quality of the starting point. Even the extremely disappointing performance of the SLTZ method on the Kehoe problem does not reflect aimless wandering of the iterates through the simplex. An equilibrium was definitely targeted; the problem was with the rate of convergence, not an actual absence of

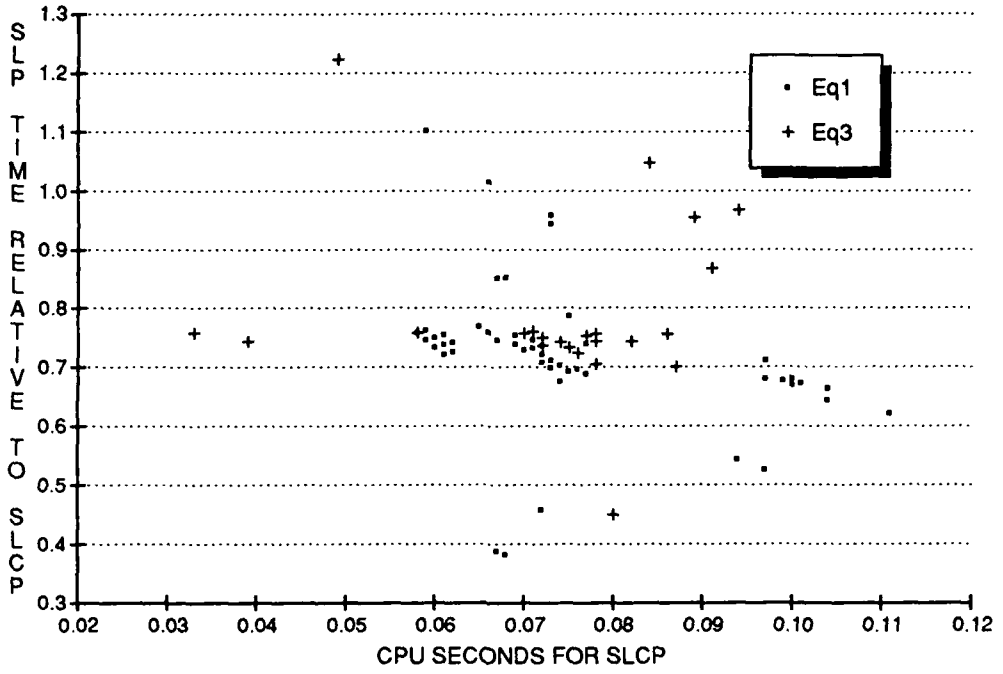

Figure 3: SLP vs. SLCP — Kehoe problem

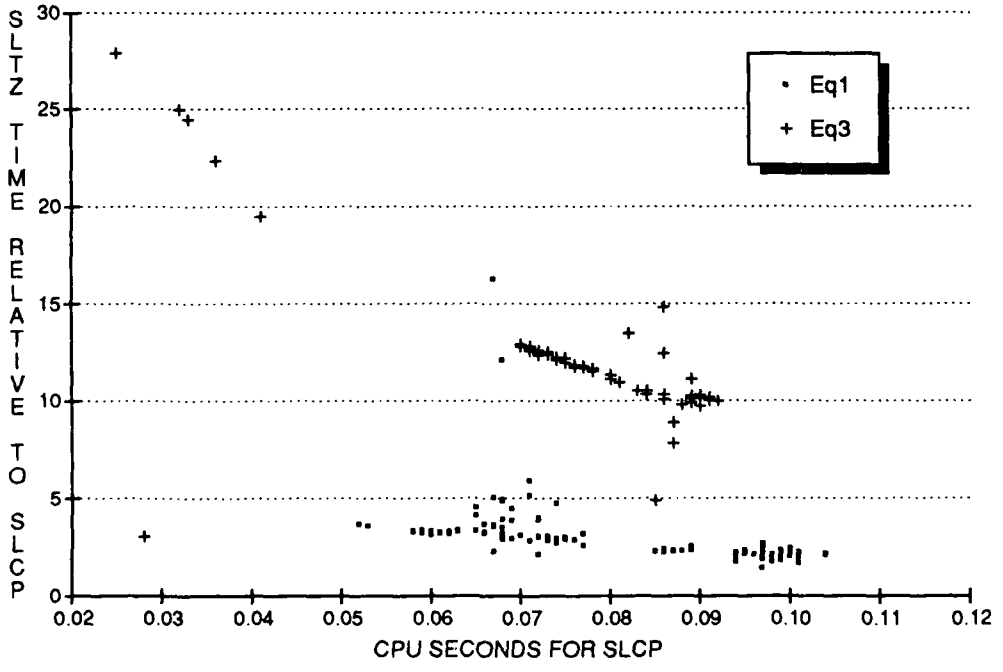

Figure 4: SLTZ vs. SLCP — Kehoe problem

convergence.

For the most part, sequential methods SLCP and SLP demonstrate comparable performance on both test problems. Solution times for SLP are on average somewhat better than those for SLCP, but there are a number of outliers for which SLP times are significantly higher. The results do not permit any conclusion as to whether supplementing SLP with a Lagrangian term or a penalty term improves overall performance. With such a mixture of results, it is probably not worth the effort to add the extra terms in practice. What may well be worth implementing is direct solution of problem Eq-DEPF. Solution times for this method rival the best of the results for SLCP and SLP, and use of Eq-DEPF does not require the user to specify a set of initial prices.

The risky venture is clearly the SLTZ method. The evidence indicates that, for a problem with a unique equilibrium, this method can compute an accurate equilibrium in as little as 10-20% of the time required by any of the other methods. On the other hand, in a multiple equilibrium context, the SLTZ method can perform so poorly as to border on being non-convergent for all practical purposes. Considerable further experimentation with other models would be required to ascertain whether it is exclusively the multiplicity characteristic that accounts for the extreme differences in the performance of SLTZ.

It is tempting to suggest the development and use of some form of hybrid method, but such hybrids are much easier to contemplate than to implement in practical software. Indeed, if a user must be prepared to "switch" to SLCP, say, when another method appears to be converging too slowly, as a practical matter it would be easier to employ SLCP throughout. If it is feasible to solve an SLCP subproblem once or a few times, it is probably feasible to solve it several times. It is often the case that the actual reason for wanting to use a lower dimensional method such as SLTZ is that the problem is too big to solve in an SLCP/SLP or Eq-DEPF context. In this case, it is of no practical value to know that SLCP or SLP would quadratically converge to the solution once we had used SLTZ to bring the iterates "close enough" to the equilibrium.

8

SUMMARY AND PERSPECTIVE

There is no question but that general equilibrium problems can be very difficult to solve. Nonetheless, the results of our research establish a number of characteristic features of CGE problems which make them in some sense more benign than the general nonlinear complementarity problem or the general (nonconvex) nonlinear program.

First, under mathematical assumptions that have reasonable economic content, the Eq-NLCP problem is known to have at least one feasible and complementary solution. In an optimization context, this means that the equivalent problem Eq-NLP is feasible and has a global minimum objective value of zero, which can in fact be attained.

Second, we have shown that, under a seemingly reasonable rank condition, a complementary (optimal) solution exists which is a vertex of the linearized constraints at that point.

Third, we have derived an easily implemented rule for constructing a price normalization which, in conjunction with a matching artificial column, guarantees that *any* linearization of the nonlinear problem has a feasible, complementary solution. Moreover, given the rank condition for existence of a basic complementary solution, this same construction guarantees that a complementary solution can be successfully computed by Lemke's method.

Fourth, we have seen that any regular equilibrium solution exerts a domain of attraction for Newton and quasi-Newton iterates. Within this domain, we may expect quadratic convergence of Newton iterates, and superlinear convergence is possible for quasi-Newton iterates. SLCP, SLP, and projected Lagrangian methods all produce Newton iterates once the complementary (optimal) basis has been determined. We have some assurance from genericity analysis that regular solutions are in fact typical in equilibrium models.

Fifth, we have derived an equivalent linearly constrained formulation of the CGE problem (Eq-DEPF) that has two important uses. One, its objective function can be used as a differentiable exact merit function for judging and perhaps guiding the iterates of any of the sequential solution methods studied. We have used this function to demonstrate that the search directions generated by any method which uses Taylor's expansions to linearize the demand functions can be considered a descent direction for some value of an arbitrary positive penalty parameter. The second contribution of this development is that optimizing problem Eq-DEPF directly can be an effective means of computing an equilibrium solution for models of an appropriate size and structure.

These are helpful and encouraging results, both in terms of improving our understanding of the computational properties of equilibrium problems and in suggesting directions for the next generation of solution procedures. There remains, nonetheless, a basic theoretical dilemma which cannot be resolved at our current level of understanding. On one hand, we have a collection of optimization methods which have been designed to produce iterates that converge from almost any starting point. Unfortunately, for the equilibrium problem, we cannot rule out the possibility that these iterates may converge to a non-equilibrium

point. On the other hand, we have Newton and other methods that obtain complementary solutions of the subproblems. These methods produce iterates that do in fact converge to an equilibrium solution, if they converge. Unfortunately, we can only suggest reasons why iterates of such methods might always be globally convergent on an equilibrium problem; nothing definitive has been established. In the face of this theoretical dilemma, however, the preponderence of the computational evidence (both our own and that of other researchers) shows that: (a) Newton-like methods do tend to produce iterates that converge from virtually any starting point, and (b) optimization approaches do tend to find the global optimum of the nonconvex equilibrium problem. This situation certainly suggests that the kinds of equilibrium models that are formulated in practice possess certain favorable computational properties that theoretical analysis has yet to discover.

Before closing this document, we first present a brief section on extending our results to equilibrium models that contain nonlinear production technologies with constant returns to scale. We then suggest what we believe to be some promising and important directions for future research.

8.1 Extension to nonlinear constant-returns production

Many computable general equilibrium models represent the production side of the economy in terms of differentiable nonlinear production or profit functions that reflect constant returns to scale. Given the theoretical and practical importance of such structures, we desire to ascertain whether or not the results of this research can be extended to apply to such models. We outline here some very recent and preliminary observations on this matter which indeed suggest that our findings have more general applicability.

Since most (and possibly all) applied general equilibrium models with nonlinear production specify an explicit profit function (or, if not, a production function for which the profit function is known), we focus our attention on models with nonlinear profit functions. In the case of constant returns to scale, a unit profit function is specified which does not determine the level of operation. Profit functions that represent closed, convex, constant-returns production possibilities are known to be continuous and homogeneous of degree one and convex in prices. For our computational purposes, it will be necessary to assume that the unit profit function is at least twice continuously differentiable. To apply the correspondence between Wilson's SQP method and SLCP, continuous third derivatives are also required. As a practical matter, most of the functions employed in applied models are infinitely differentiable wherever they are once differentiable.

We generalize the notion of n linear production activities to n independent (constant-returns) production sectors whose operations are represented by unit profit functions $r_j(p)$. Let $r(p)$ be the n-vector of these individual unit profit functions. Since $r(p)$ is homogeneous of degree one, Euler's law implies that $r(p) = \nabla r(p)p$. It follows that the excess profit conditions, $r(p) \leq 0$, may be equivalently stated as $\nabla r(p)p \leq 0$. The gradient functions are in turn homogeneous of degree zero, which implies that $\nabla^2 r_j(p)p = 0$ for all j. The Hessians $\nabla^2 r_j(p)$ are, of course, symmetric and are furthermore positive semidefinite by

virtue of the convexity of the $r_j(p)$. By the lemma of Hotelling, the profit-maximizing net-output combination (per unit of operation) is given by $[\nabla r_j(p)]^T$. A positive component of this vector represents a net output of the commodity; a negative component indicates a net input.

We again follow the presentations of Kehoe [Keh 82] and Mas-Colell [MasC 85]. To mimic our earlier notation for the linear production case, we define a matrix function $A(p) \equiv \nabla^T r(p)$. The levels of operation for the individual production sectors are represented by the n-vector y. For the appropriate analogies to properties (A1) and (A2) of Section 2.2, we reasonably assume the availability of free disposal and that no prices exist for which $A(p)$ permits positive output without input. Existence of equilibrium then follows from the usual fixed-point arguments; see either of the above references.

Given existence, we may now specify the equilibrium problem as the following specially structured NLCP:

Find a complementary solution of:

$$\begin{array}{lrcl} \text{(CR1)} & -d(p) + A(p)y + hv & \geq 0 \perp & p \geq 0 \\ \text{(CR2)} & -A^T(p)p & \geq 0 \perp & y \geq 0 \\ \text{(CR3)} & -h^T p & = -1 \perp & v \\ \text{(CR4)} & p \quad , \quad y & \geq 0 & \end{array}$$

Note that this formulation has the same basic structure as problem Eq-NLCP. The only difference is that the previously constant matrix A has been replaced by the matrix function $A(p)$. This generalization in no way affects the analysis used in Section 3.3 to demonstrate that the problem can be equivalently stated as a nonlinear program which minimizes v over the same inequalities. Similarly, a formulation analogous to Eq-DEPF can be constructed, although, in light of the additional nonlinearity of $A(p)$, this formulation will more resemble the general case (DEPF) than the special case Eq-DEPF (see Section 5.2.1). Thus, we may also apply the theory, algorithms, and convergence analysis of optimization methods to the equilibrium problem with nonlinear, constant-returns production. In particular, given adequate differentiability of the unit profit functions, the basic equivalence between SLCP and Wilson's SQP method applies to this problem as well. What we have not yet investigated is the effect of the nonlinear $A(p)$ on the descent results we derived in Chapter 5.

In applying Newton's method to the above NLCP, the linearized subproblem which is constructed at a linearization point (p^k, y^k) is the following:

Find a complementary solution of:

$$\begin{array}{lrcccccl} \text{(LCR1)} & \left[\sum_j y_j^k \nabla^2 r_j(p^k) - \nabla d(p^k)\right] p & + & A(p^k)y & + & hv & \geq & d(p^k) \quad \perp \quad p \geq 0 \\ \text{(LCR2)} & -A^{\mathsf{T}}(p^k)p & & & & & \geq & 0 \quad \perp \quad y \geq 0 \\ \text{(LCR3)} & -h^{\mathsf{T}}p & & & & & = & -1 \quad \perp \quad v \\ \text{(LCR4)} & p \quad , & & y & & & \geq & 0 \end{array}$$

The above relations are also the appropriate linearized constraints for applying any of the optimization methods we have discussed in this research. Two observations are important.

First, the constraint matrix indeed has the form we studied in Chapter 6. This means that, if the relations (LCR1)–(LCR4) are consistent, the subproblem has a complementary solution which can be successfully computed by Lemke's method. Given the assumption that $A(p)$ satisfies property (A2) of Section 2.2 for all p, we can in fact conclude that each subproblem is feasible, provided that the price normalization is constructed as prescribed in Chapter 6.

Second, the symmetry and positive semidefiniteness of the Hessian terms appearing in the above linearization may in fact cause the subproblem to be better behaved than the subproblems encountered with linear production. In particular, solving the subproblem as a quadratic program may be more reliable in this case. Moreover, it is intuitive to suspect that the submatrix $\left[\sum_j y_j^k \nabla^2 r_j(p^k) - \nabla d(p^k)\right]$ is more amenable to approximation by a symmetric positive semidefinite matrix, such as we studied with the SLTZ method in the previous chapter.

It would appear, then, that most of the results we have obtained in this research will also apply to the important case of nonlinear production with constant returns to scale, given the existence of sufficiently differentiable unit profit functions. Further analysis is required to ensure that we have not overlooked any subtleties arising in the nonlinear situation. If not, this extension will greatly broaden the applicability of our research.

8.2 Future research

We believe that it is extremely important to better understand the empirical observations that Newton methods (and other linearization methods) tend to be globally convergent when applied to problem Eq-NLCP and that optimization algorithms successfully find the global minimum of nonconvex problems Eq-NLP and Eq-DEPF. We think there is more to be learned in this regard from study of the optimality conditions for the optimization formulations of the equilibrium problem. In particular, given properties (H2) and (W2) of Section 2.1, further investigation of the second-order conditions may be fruitful, even though our analysis to date has not revealed anything definitive. Since many of the models solved in practice employ sums of demand functions derived from Cobb-Douglas or more general constant-elasticity-of-substitution utility functions, it may well be that special properties

of these functions are the source of good computational performance. This suggests also studying the optimality conditions as specialized to these popular functional forms.

We think it worthwhile to develop and evaluate computational procedures which combine any (and all) of the sequential methods studied with a rigorous line search and choice of step length based on the merit function available through problem Eq-DEPF. The principal intent here is both to avoid uphill steps (such as have been seen to occur with the SLTZ method, for example) and to stabilize the initial iterations of the algorithms. It is easy to focus on asymptotic behavior of the iterates, but we have observed repeatedly in our computational experiments that the first few linearizations can be critical in terms of establishing the path that is to be followed to the equilibrium solution. It is here that a line search accounting for the divergence of the linearization from the true functions could prevent wasted iterations on extreme and unrealistic subproblem solutions.

A major disappointment of the current research concerns the equivocal results obtained for symmetric linearization approaches (such as SLTZ) that allow for the solution of a subproblem in a lower dimension. Not only does the theoretical analysis of Chapter 5 demonstrate that such methods can produce uphill search directions, the computational results of Chapter 7 confirm that such behavior does in fact occur on actual models. On the other hand, in some contexts the symmetric methods could perform many times better than the other methods studied. It is a true irony that we set out in this research to establish the viability of symmetric approximation methods, but we seem to have done more to demonstrate their inherent riskiness.

We see two possible approaches to investigate for models of a size that makes directly solving Eq-DEPF or repeatedly solving SLCP/SLP subproblems a costly undertaking. The first avoids the solution of such subproblems by using symmetric linearizations. It is then necessary to devise (somehow) an overall sequential scheme that is both globally convergent and convergent at an acceptable (linear) rate. The second formulates subproblems based on Taylor's expansions (thereby promoting overall convergence at a quadratic rate) but seeks alternative procedures for solving the (large) subproblems that may be less costly than procedures known to date.

For the first approach, we may consider using a rigorous Eq-DEPF line search to supervise a symmetric linearization procedure (whether it be SLTZ, a Jacobi method, or some form of projection method). This certainly would prevent the kind of uphill excursion that we observed on Kehoe's multiple equilibrium problem. Unfortunately, it is not at all clear what to do if the subproblem solution provides a search direction that is uphill for all positive step lengths. This is where further research may suggest a superior formulation of the subproblem. Given the availability now of the Eq-DEPF merit function, we are in a much better position to pursue such research than previously.

The other direction is seeking improved methods for solving the specially structured subproblems which arise in the course of SLCP or SLP. Current solution procedures based on Lemke's method or MINOS, for instance, can take advantage of the general sparsity of the matrix of linearized constraints, but they cannot make any use of the skew-symmetric

structure corresponding to the production activities. This suggests the study and development of special factorization routines that recognize the special structure of complementary and almost-complementary bases for these subproblems. The block-angular nature of the subproblem also makes it amenable to solution by a decomposition procedure, particularly in the linear case of an SLP subproblem or the convex case of an SQP subproblem (which utilizes the Jacobian of the demand functions in the constraints but uses a positive semidefinite approximation in the objective function). Earlier thoughts on applying decomposition to linearized equilibrium problems are presented at some length in [Sto 85]. An effective decomposition strategy would lead to solving subproblems that are roughly the same size as those encountered in using a symmetric linearization method.

On balance, we believe that we have made some valuable incremental progress in this research, but much remains to be studied and understood before the solution of large-scale (and structurally general) general equilibrium models becomes a routine undertaking.

BIBLIOGRAPHY

[AH 82] B.-H. Ahn and W.W. Hogan, On convergence of the PIES algorithm for computing equilibria, *Operations Research* 30 (1982) 281–300.

[Avr 76] M. Avriel, *Nonlinear Programming: Analysis and Methods* (Prentice-Hall, Englewood Cliffs, New Jersey, 1976).

[Bro 85] M.N. Broadie, An introduction to the octahedral algorithm for the computation of economic equilibria, *Mathematical Programming Study* 23 (1985) 121–143.

[Cot 64] R.W Cottle, Note on a fundamental theorem in quadratic programming, *Journal of the Society for Industrial and Applied Mathematics* 12 (1964) 663–665.

[Cot 76] R.W. Cottle, Complementarity and variational problems, *Symposia Mathematica* XIX (1976) 177–208.

[DEG 79] G.B. Dantzig, B.C. Eaves and D. Gale, An algorithm for a piecewise linear model of trade and production with negative prices and bankruptcy, *Mathematical Programming* 16 (1979) 190–209.

[DN 84] S. Dafermos and A. Nagurney, A network formulation of market equilibrium problems and variational inequalities, *Operations Research Letters* 3 (1984) 247–250.

[Eav 71] B.C. Eaves, The linear complementarity problem, *Management Science* 17 (1971) 612–634.

[Eav 78] B.C. Eaves, A locally quadratically convergent algorithm for computing stationary points, Technical Report SOL 78-13, Systems Optimization Laboratory, Department of Operations Research, Stanford University (May 1978).

[Eav 87] B.C. Eaves, Thoughts on computing market equilibrium with SLCP, in A.J.J. Talman and G. van der Laan (eds.), *The Computation and Modeling of Economic Equilibria* (North-Holland, Amsterdam, 1987) 1–17.

[Gar 73] C.B. Garcia, Some classes of matrices in linear complementarity theory, *Mathematical Programming* 5 (1973) 299–310.

[GMSW 86] P.E. Gill, W. Murray, M.A. Saunders and M.H. Wright, Some theoretical properties of an augmented Lagrangian merit function, Technical Report SOL 86-6, Systems Optimization Laboratory, Department of Operations Research, Stanford University (September 1986).

[GMSW 87] P.E. Gill, W. Murray, M.A. Saunders and M.H. Wright, Constrained nonlinear programming, Technical Report SOL 87-13, Systems Optimization Laboratory, Department of Operations Research, Stanford University (December 1987), to appear in G.L. Nemhauser, A.H.G. Rinnooy Kan and M.J. Todd (eds.), *Handbooks in Operations Research and Management Science–Volume 1, Optimization* (North-Holland, Amsterdam, 1988).

[GMW 81] P.E. Gill, W. Murray and M.H. Wright, *Practical optimization* (Academic Press, London, 1981).

[GS 61] R.E. Griffith and R.A. Stewart, A nonlinear programming technique for the optimization of continuous processing systems, *Management Science* 7 (1961) 379–392.

[GW 81] V.A. Ginsburgh and J.L. Waelbroeck, *Activity analysis and general equilibrium modeling* (North-Holland, Amsterdam, 1981).

[Jo-N 79] N.H. Josephy, Newton's method for generalized equations, Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin-Madison (1979), available from National Technical Information Service, accession number AD A077 096.

[Jo-Q 79] N.H. Josephy, Quasi-Newton methods for generalized equations, Technical Summary Report 1966, Mathematics Research Center, University of Wisconsin-Madison (1979), available from National Technical Information Service, accession number AD A077 097.

[Keh 82] T.J. Kehoe, Regular production economies, *Journal of Mathematical Economics* 10 (1982) 147–176.

[Keh 84] T.J. Kehoe, Computing all of the equilibria in economics with two factors of production, *Journal of Mathematical Economics* 13 (1984) 207–223.

[Keh 85] T.J. Kehoe, A numerical investigation of multiplicity of equilibria, *Mathematical Programming Study* 23 (1985) 240–258.

[Kyp 86] J. Kyparisis, Uniqueness and differentiability of solutions of parametric nonlinear complementarity problems, *Mathematical Programming* 36 (1986) 105–113.

[Lem 68] C.E. Lemke, Bimatrix equilibrium points and mathematical programming, *Management Science* 11 (1965) 681–689.

[Mang 80] O.L. Mangasarian, Locally unique solutions of quadratic programs, linear and nonlinear complementarity problems, *Mathematical Programming* 19 (1980) 200–212.

[Mann 85] A.S. Manne (ed.), Economic equilibrium: model formulation and solution, *Mathematical Programming Study* 23 (1985).

[MasC 85] A. Mas-Colell, *The theory of general economic equilibrium: a differentiable approach* (Cambridge University Press, Cambridge, England, 1985).

[Mat 87] L. Mathiesen, An algorithm based on a sequence of linear complementarity problems applied to a Walrasian equilibrium model: an example, *Mathematical Programming* 37 (1987) 1–18.

[MCW 80] A.S. Manne, H.-P. Chao and R. Wilson, Computation of competitive equilibria by a sequence of linear programs, *Econometrica* 48 (1980) 1595–1615.

[MR 83] L. Mathiesen and T.F. Rutherford, Testing the robustness of an iterative LCP algorithm for solving Walrasian equilibrium models, Discussion Paper 0883, Norwegian School of Economics and Business Administration (1983).

[MS 82] B.A. Murtagh and M.A. Saunders, A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints, *Mathematical Programming Study* 16 (1982) 84–117.

[Mur 83] F.H. Murphy, Design strategies for energy market models, in B. Lev (ed.), *Energy models and studies* (North-Holland, Amsterdam, 1983) 45–64.

[Nag 87] A. Nagurney, Competitive equilibrium problems, variational inequalities and regional science, *Journal of Regional Science* 27 (1987) 503–517.

[PC 82] J.S. Pang and D. Chan, Iterative methods for variational and complementarity problems, *Mathematical Programming* 24 (1982) 284–313.

[Phi 85] R.L. Phillips, Computing solutions to generalized equilibrium models by successive under-relaxation, *Mathematical Programming Study* 23 (1985) 192–209.

[RK 72] J.B. Rosen and J. Kreuser, A gradient projection algorithm for nonlinear constraints, in F.A. Lootsma (ed.), *Numerical Methods for Non-Linear Optimization* (Academic Press, London and New York, 1972) 297–300.

[Rob 72] S.M. Robinson, A quadratically convergent algorithm for general nonlinear programming problems, *Mathematical Programming* 3 (1972) 145–156.

[Rob 74] S.M. Robinson, Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms, *Mathematical Programming* 7 (1974) 1–16.

[Rob 80] S.M. Robinson, Strongly regular generalized equations, *Mathematics of Operations Research* 5 (1980) 43–62.

[Rob 82] S.M. Robinson, Generalized equations, in A. Bachem, M. Grotschel and B. Korte (eds.), *Mathematical Programming: The State of the Art* (Springer-Verlag, Berlin, 1982) 346–367.

[Rut 86] T.F. Rutherford, *Applied general equilibrium modeling*, Doctoral Dissertation, Department of Operations Research, Stanford University (1986).

[Sca 73] H.E. Scarf, *The computation of economic equilibria* (Yale University Press, New Haven, 1973).

[SMD 87] J.C. Stone, P.H. McAllister and G.B. Dantzig, Using the PILOT model to study the effects of technological change, in B. Lev, J.A. Bloom, A.S. Gleit, F.H. Murphy and C. Shoemaker (eds.), *Strategic Planning in Energy and Natural Resources* (North-Holland, Amsterdam, 1987) 31–42.

[Sto 85] J.C. Stone, Sequential optimization and complementarity techniques for computing economic equilibria, Technical Report SOL 84-9, Systems Optimization Laboratory, Department of Operations Research, Stanford University (November 1984), reproduced in part in *Mathematical Programming Study* 23 (1985) 173–191.

[SW 73] J.B. Shoven and J. Whalley, General equilibrium with taxes: a computational procedure and an existence proof, *Review of Economic Studies* 60 (1973) 475-490.

[Wil 63] R.B. Wilson, *A simplicial algorithm for concave programming*, Doctoral Dissertation, Graduate School of Business Administration, Harvard University (1963).

[ZKL 85] J. Zhang, N.-H. Kim and L. Lasdon, An improved successive linear programming algorithm, *Management Science* 31 (1985) 1312–1331.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR # SOL 88-7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Formulation and Solution of Economic Equilibrium Problems		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John C. Stone		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0343
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research - SOL Stanford University Stanford, CA 94305-4022		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 1111MA
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research - Dept. of the Navy 800 N. Quincy Street Arlington, VA 22217		12. REPORT DATE April 1988 13. NUMBER OF PAGES pp. 77
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

economic equilibrium
complementarity problems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Please see back of form.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TR SOL 88-7: Formulation and Solution of Economic Equilibrium Problems, by John S. Stone

We develop and assess a number of equivalent mathematical formulations of the general equilibrium problem in economics. We begin with the traditional representation as a nonlinear complementarity problem and develop alternative representations as nonlinear optimization problems. All of our formulations depart from previous approaches by including an explicit linear equality for price normalization and a matched artificial variable which must be zero at any equilibrium solution. This structure has the theoretical and computational advantage that any linearization of the equilibrium problem has a feasible complementary solution. Moreover, under a reasonable rank condition, a basic complementary solution exists which can be successfully computed by Lemke's almost-complementary pivoting method.

We describe five general-purpose methods which can be applied to solving the equilibrium problem. The common feature of these methods is solving a sequence of linearized problems. We establish a number of equivalences between the methods, when applied to the equilibrium problem, and assess their local and global convergence properties in that context. An important new tool in this analysis is another problem formulation based on a differentiable exact penalty function. This formulation provides, perhaps for the first time, a rigorous means of evaluating the progress of a sequential method for computing an equilibrium solution. Our analysis reveals a basic theoretical dilemma in solving general equilibrium problems by these sequential methods. One group of methods produces iterates that converge from any starting point, but the sequence may converge to a non-equilibrium point. Another group produces iterates that may fail to converge, but successfully converging sequences do attain an equilibrium.

We perform a number of computational experiments on two small problems from the literature. The results show considerable variation in the solution times for the various methods, but all methods succeed in locating an equilibrium, even from poor starting points. This successful performance (in addition to that reported by other researchers) suggests that the kinds of general equilibrium models formulated in practice possess certain favorable computational properties that theoretical analysis has yet to discover.